GENOMIC STUDIES OF SOCIAL EVOLUTION IN BEES

BY

SARAH HOLLIS WOODARD

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Biology
with a concentration in Ecology and Evolutionary Biology
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2012

Urbana, Illinois

Doctoral Committee:

Professor Gene E. Robinson, Chair, Director of Research
Professor Sydney A. Cameron
Professor Matthew E. Hudson
Professor Hugh M. Robertson

ABSTRACT

The eusocial insects (ants, termites, and some bees and wasps) are paragons of social complexity, with their highly organized societies comprised of primarily sterile workers, headed by one or a few reproductives. Despite a rich body of theoretical, empirical, and natural history research on social insects, relatively little is known about how the insect societies evolved at the molecular level. Here, I employ techniques from genomic biology to shed light on the molecular basis of social evolution in the bees, a group of socially diverse species that encompasses multiple, independent evolutions of eusociality. In Chapter 1, I provide a detailed overview of the chapters contained in this dissertation. In Chapter 2, I review current progress in molecular evolutionary analyses of the social insects and outline some common themes emerging from this area of research. In Chapter 3, I use a comparative genomics approach to identify hundreds of genes that are rapidly evolving in eusocial bees. Additionally, I highlight biological processes, such as carbohydrate metabolism, which may have been important targets of natural selection during the evolution of eusociality in these bee lineages. In Chapter 4, I combine a behavioral analysis with a microarray experiment to explore how changes in maternal traits in queens of the bumble bee *Bombus terrestris* are regulated by workers during the nest initiation phase of the *Bombus* life cycle. Here, I provide evidence that workers socially regulate both brood-feeding and egg-laying behavior in queens, and may also influence patterns of brain gene expression in queens. Lastly, in Chapter 5, I use a microarray experiment on *B. terrestris* to identify patterns of brain gene expression that are shared by brood-feeding bumble bee queens and workers, as well as genes that appear to be associated with brood care in bumble bees and in other eusocial insect lineages. These shared patterns of brain gene expression may provide insights into how sibling care evolved in the bumble bees, as well as how convergent evolution of cooperative brood care occurred across multiple eusocial insect lineages.

ACKNOWLEDGMENTS

I am very grateful to my advisor, Gene Robinson, for providing me with more support, inspiration, and opportunities over the course of my Ph.D. than I could have imagined possible. I owe special thanks to my labmate and best friend Brielle Fischman, who was co-first author on two manuscripts included in this thesis, and with whom I share the fondest memories from my Ph.D. Hugh Robertson, Matthew Hudson, and Sydney Cameron served on my committee and provided me with invaluable input and guidance, and I consider them colleagues and friends. For the work presented in Chapter 3, I would like to especially thank Aarti Venkat, Kranthi Varala, and Andrew Clark at Cornell University, without whom the project would not have been possible. For the work presented in Chapters 4 and 5, I would like to especially thank Guy Bloch, Hagai Shpigler, and other members of the Bloch lab for graciously hosting me at the Hebrew University of Jerusalem, Israel, and also Mark Band and Jenny Drnevich of the Keck Center, University of Illinois at Urbana-Champaign. I would like to thank all of my fellow Robinson lab members and friends for their input and assistance, in particular Amy Toth, Tim Daugherty, Tara McGill, Claudia Lutz, Seth Ament, Karen Kapheim, and Tom Newman. I would also like to thank Chase Howell for his friendship and culinary prowess. Finally, my dissertation research was made possible by funding from the National Science Foundation, the National Institutes of Health, and from my program, the Program in Ecology, Evolution, and Conservation at the University of Illinois at Urbana-Champaign.

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION

The eusocial insects (ants, termites, and some bees and wasps) exhibit some of the most extreme levels of social complexity found in the Animal Kingdom, with their large, highly organized societies comprised of mostly sterile workers, headed by one or a few reproductives (1). Eusocial insects have long served as a source of fascination for biologists because at first glance, they represent a paradox; how can selection favor the evolution and maintenance of sterile individuals, despite strong selective pressure for individual reproductive success (2)? For more than a hundred years, this question, together with additional questions such as how intranest cohesion is facilitated, and what similarities exist between the eusocial insects and other group-living organisms, has inspired a rich and diverse body of research on the eusocial insects (1, 3).

Our ability to explore the evolution of the insect societies expanded considerably with the rise of genomic biology. Beginning with the publication of the first fully sequenced eusocial insect genome, that of the honey bee *Apis mellifera* in 2006 (4), followed by the publication of several sequenced ant genomes (5-9) and the release of Expressed Sequence Tag (EST) sets for numerous other species (10-14), we have learned a great deal about the biology of eusocial insect genomes. In addition to providing descriptive information about the content of eusocial insect genomes, these genomic datasets have been used to test various molecular predictions of long-standing hypotheses in social insect biology (13, 15).

This dissertation uses multiple techniques from genomic biology to explore the molecular basis of social evolution in the insects, with an emphasis on bees, a group that includes species that span the continuum of insect social complexity. In Chapter 2, I review the current state of molecular evolutionary research on the eusocial insects. In addition to describing some of the most prominent insights that

have emerged from this field, I also provide a synthesis wherein I outline some common themes and discuss potential challenges and important directions for the future of the field.

In Chapter 3, I apply a comparative genomics approach to EST data from nine bee species, combined with data from the honey bee (*A. mellifera*) genome, to search for genes that are rapidly evolving across multiple, independent evolutions of eusocial bees. The presence of rapidly evolving sites among protein sequences can be indicative of the action of natural selection, and the detection of these "signatures of selection" across multiple lineages in a phylogeny can indicate that similar selective pressures have been shared by lineages (16). Using this approach, I identified hundreds of genes out of a set of ~3,600 orthologous sequences that appear to be under different selective pressures in eusocial and non-eusocial bee lineages, including many genes that are rapidly evolving within a single eusocial lineage, and many that are rapidly evolving in multiple eusocial lineages.

In Chapter 4, I combine detailed behavioral observations with microarrays in a social manipulation experiment designed to study social regulation of maternal traits in nest-founding bumble bee queens. During the nest-founding phase of the colony cycle, mated queens emerge, initiate nests, and feed their offspring, but upon the emergence of the first workers in the colony, queens cease feeding brood and instead become specialized primarily on egg-laying behavior. Here, I provide evidence that workers socially regulate this behavioral transition in nest-founding queens, and that this plasticity is bidirectional; queens can greatly reduce their brood-feeding behavior and increase egg-laying if workers are prematurely added to their nests, and alternatively, they can continue to feed brood and reduce egg-laying if workers are artificially removed from their nests. Using Agilent EST-based microarrays (14), I used brain gene expression data to show that workers do appear to have an effect on brain gene expression in these queens, but these social effects are much stronger in queens later in founding phase, who have had workers emerge in their nests.

In Chapter 5, I use a microarray experiment to explore the hypothesis that sibling care evolved from maternal care in bumble bees by testing the prediction that nest-founding queens and brood-feeding workers, who exhibit maternal and sibling care, respectively, have similar patterns of brain expression for genes related to feeding and reproduction, relative to other individuals in the nest. This experiment was modeled after Toth *et al.*'s tests of this prediction in *Polistes* (13, 15), and I draw comparisons between my data, Toth *et al.*'s (13, 15), and data from a highly eusocial species, the honey bee *Apis mellifera* (17, 18), to explore a second hypothesis, that there are genetic "toolkits" related to feeding and reproduction that are conserved across solitary and social insects, across independent evolutions of eusociality (19, 20).

Recent advances in genomic biology have revolutionized our capacity to explore the evolution of social complexity in the eusocial insects. In this dissertation, I discuss my own contributions to this exciting area of research, which used numerous techniques and methodologies drawn from the fields of genomic and molecular biology, evolutionary biology, neuroscience, and animal behavior, and which were guided by both long-standing and more recently-developed hypotheses in social insect biology. The research contained herein was inspired by the desire to understand how and why sociality evolved in the eusocial insects, because these fascinating creatures are models for how evolution can give rise to complex systems in which individuals cooperate in extraordinary ways.

LITERATURE CITED

1. Wilson EO (1971) *The Insect Societies* (Belknap Press, Cambridge, MS).
2. Darwin, C*. The Origin of Species by Means of Natural Selection* (John Murry, London, 1859).
3. Wilson, EO (1975) *Sociobiology: the New Synthesis.* Belknap Press of Harvard University Press, Cambridge, MA.
4. HBGSC (2006) Insights into social insects from the genome of the honeybee *Apis mellifera*.

Nature 443:931-949.

5. Bonasio R, et al. (2010) Genomic comparison of the ants *Camponotus floridanus* and *Harpegnathos saltator*. *Science* 329:1068-1071.
6. Nygaard S, *et al.* (2011) The genome of the leaf-cutting ant *Acromyrmex echinatior* suggests key adaptations to advanced social life and fungus farming. Genome Res 21:1339-1348.
7. Smith CR, et al. (2011) Draft genome of the red harvester ant *Pogonomyrmex barbatus. Proc Natl Acad Sci USA.*
8. Smith CD, et al. (2011) Draft genome of the globally widespread and invasive Argentine ant (*Linepithema humile*). *Proc Natl Acad Sci USA.*
9. Wurm Y, et al. (2011) The genome of the fire ant *Solenopsis invicta*. *Proc Natl Acad Sci USA.*
10. Hoffman EA, Goodisman MA (2007) Gene expression and the evolution of phenotypic diversity in social wasps. *BMC Biol.* 5:23.
11. Sadd BM, Kube M, Klages S, Reinhardt R, Schmid-Hempel P (2010) Analysis of a normalised expressed sequence tag (EST) library from a key pollinator, the bumblebee *Bombus terrestris*. *BMC Genomics* 11:110.
12. Steller MM, Kambhampati S, Caragea D (2010) Comparative analysis of expressed sequence tags from three castes and two life stages of the termite *Reticulitermes flavipes*. *BMC Genomics* 11:463.
13. Toth AL, et al. (2007) Wasp gene expression supports an evolutionary link between maternal behavior and eusociality. *Science* 318: 441-444.
14. Woodard SH, et al. (2010) Genes involved in convergent evolution of eusociality in bees. *Proc Natl Acad Sci USA* 7472-7477.
15. Toth AL, Varala K, Henshaw MT, Rodriguez-Zas SL, Hudson ME, Robinson GE (2010). Brain transcriptomic analysis in paper wasps identifies genes associated with behaviour across social insect lineages. *Proc Roy Soc Lond B* 277:2139-2148.

16. Yang ZH, Bielawski JP (2000) Statistical methods for detecting molecular adaptation. *Trends Ecol Evol* 15:496–503.
17. Whitfield CW, Cziko AM, Robinson GE (2003) Gene expression profiles in the brain predict behaviour in individual honey bees. *Science* 302:296–299.
18. Alaux C, *et al*. (2009). Honey bee aggression supports a link between gene regulation and behavioral evolution. *Proc Natl Acad Sci USA* 106:15400-15405.
19. Smith CR, Toth AL, Suarez AV, Robinson GE (2008) Genetic and genomic analyses of the division of labour in insect societies. *Nat Rev Genet* 9:735-748.
20. Toth AL, Robinson GE (2007) Evo-devo and the evolution of social behavior. *Trends Genet* 23:334-341.

CHAPTER 2: MOLECULAR EVOLUTIONARY ANALYSES OF INSECT SOCIETY: PROGRESS AND PROSPECTS[1]

Abstract

The social insects live in extraordinarily complex and cohesive societies, where many individuals sacrifice their personal reproduction to become helpers in the colony. Identifying adaptive molecular changes involved in eusocial evolution in insects is important for understanding the mechanisms underlying transitions from solitary to social living, and the maintenance and elaboration of social life. Here I review recent advances made in this area of research in several insect groups, the ants, bees, wasps, and termites. Drawing from whole genome comparisons, candidate gene approaches, and a genome-scale, comparative analysis of protein-coding sequence, I highlight novel insights gained for five major biological processes; chemical signaling, brain development and function, immunity, reproduction, and metabolism and nutrition. Lastly, I compare across these diverse approaches and social insect lineages, and discuss potential common themes of eusocial evolution, and challenges and prospects for future research in the field.

---

[1] Chapter 2 was published in *Proceedings of the National Academy of Sciences* in 2011: Fischman BJ*, Woodard SH*, Robinson GE (2011) Molecular Evolutionary Analyses of Insect Society: Progress and Prospects. *Proc. Natl Acad. Sci. USA* 108: 10847-54.

[*] These authors contributed equally to this work

Author contributions: B.J.F., S.H.W., and G.E.R. wrote the paper.

The social insects are exemplars of cooperative group living. Within their complex societies, there is a reproductive division of labor in which only a small number of individuals reproduce, while all other individuals belong to a functionally sterile worker caste that specializes on tasks important for colony growth and development (1). Whereas there has been much theoretical research on the evolutionary forces that may select for eusociality (2, 3), less is known about the actual molecular mechanisms involved in transitions from solitary to social living and in the maintenance and elaboration of eusociality in insects (4).

The social insects provide a powerful comparative framework for investigating mechanisms involved in eusocial evolution. Eusociality has arisen independently at least 12 times in the insects (5-8) and social insects have all converged on the following three characteristics: reproductive division of labor, cooperative brood care, and overlapping generations (9). Additionally, despite sharing this core set of traits, there are many differences among eusocial lifestyles, which may be related to ecological, phylogenetic, or other factors specific to particular eusocial lineages (1). By comparing across social insect lineages, it is possible to both search for common mechanisms of eusocial evolution and explore how eusociality evolves under different conditions.

Analysis of adaptive evolution at the molecular level can yield great insights into the mechanisms underlying the evolution of complex phenotypes, such as eusociality. Genomic sequence provides a molecular record of how natural selection has shaped an organism's evolutionary history (10). Several methods have been developed for comparing genes and genomes to identify molecular signatures of adaptation. These methods were largely developed during the pre-genomic era (11), but gain enormous power when large genomic datasets are available, particularly for sets of closely related, phenotypically variable species (12, 13). For example, comparisons of primate genomes have identified adaptive genetic changes involved in the evolution of brain size in humans (14), and comparisons of

Drosophilid genomes have shed light on the ecological pressures that shaped speciation in this group (13).

Here I review some of the first contributions of molecular evolutionary research to our understanding of eusocial evolution in insects. This research has focused on the most well-studied social insects, which include several eusocial lineages within the order Hymenoptera, the ants, bees, and wasps, and the one eusocial lineage in the order Blattodea, the termites (Fig. 1). Some studies have performed targeted, molecular evolutionary analyses of candidate genes which have been particularly valuable in species for which large amounts of genomic sequence are not yet available. Others have focused on comparative analyses of whole genome sequence, currently available for six social insects, the honey bee *Apis mellifera* (15) plus five ant species (16-19), and many solitary insects, including three solitary hymenopterans in the parasitoid jewel wasp genus *Nasonia* (20).

I also draw heavily from our own recent, genome-scale study of protein-coding sequence evolution in bees ("bee molecular evolution study"). This study analyzed ~3600 genes from a set of 10 social and non-social bee species, which encompass three independent origins of eusociality (21). Hundreds of genes were identified that exhibit a molecular signature of rapid evolution associated with sociality, defined as a higher ratio of nonsynonymous to synonymous nucleotide substitutions (d_N/d_S) in social relative to non-social bee lineages (21). Throughout this review, evidence for rapid evolution is based on relative d_N/d_S and positive selection is defined as $d_N/d_S > 1$, unless otherwise specified.

Genes identified in these studies are listed in Table 1. The insights gained from these studies have implications for understanding how evolutionary changes in the following five major biological processes might be involved in the evolution of eusociality: chemical signaling, brain development and function, immunity, reproduction, and metabolism and nutrition. I discuss evidence and predictions for the putative functional effects of identified molecular changes in these processes on social phenotypes. I also speculate on the potential adaptive significance of these molecular changes and, consider whether

these changes evolved in response to the origin, the maintenance, or the elaboration of eusociality, as each case likely involved a distinct set of selective forces. For the purposes of interpreting and synthesizing results across multiple studies I present each process separately, but it is important to recognize that these biological processes may evolve in concert and some molecular changes could potentially affect multiple processes. I end with a discussion of future prospects and challenges for this young field.

Chemical signaling

Social insects use pheromones to coordinate the behavior and physiology of colony members, such as directing the foraging activity of nestmates, reinforcing dominance status, and inhibiting ovary development in workers (45). It is unknown whether chemical signaling was important during the origins of eusociality, as other mechanisms to mediate social interactions, such as physical interactions, serve similar functions in some social insect societies (1). However, chemical signaling is certainly involved in the maintenance and elaboration of eusociality, as it is crucial for the coordination and control of colony members. In humans, where vocalization is a major component of social communication, molecular signatures of adaptation have been detected in genes underlying both the production (46) and perception (13) of vocal signals. Early studies in social insects suggest that analogous changes have occurred in the molecular machinery underlying the production and perception of chemical signals.

Gland development. Our bee molecular evolution study identified ~200 genes evolving more rapidly in social relative to non-social bee lineages (21). Gene Ontology enrichment analysis revealed that this set of genes was enriched for genes involved in gland development. This supports a role for these genes in chemical signaling, as glands are the primary organs involved in pheromone production in insects.

Moreover, the evolution of complex chemical signaling in the social insects has been associated with the diversification of the gland repertoire (1).

In other organisms, modular evolution, in which semi-autonomous genetic pathways evolve as a functional unit and are re-used in multiple contexts, appears to be a common evolutionary mechanism involved in morphological diversification (47). The sequence changes identified in genes involved in gland development in social bees may have caused modular changes to the gland development program resulting in functional changes to existing glands or the appearance of entirely new glands. This is supported by the evidence that several of these genes (*decapentaplegic, thickveins, PDGF- and VEGF-related factor 1*) have specific roles in gland patterning during early development in *Drosophila* (22, 23).

As diversification of gland function is a common characteristic shared by all social insects, it would be fruitful to investigate the sequence evolution and function of these genes in other social insect groups. It is possible that molecular changes in the same or similar genes were involved in gland evolution across other independent eusocial lineages.

Odorant receptors. Given the diversity of chemical signals used by social insects, odorant receptor genes (ORs) have been predicted to be important targets of selection during eusocial evolution (48). Early support for this prediction was found in the genome of the honey bee, *A. mellifera*, which at the time of its publication, contained the largest number of OR genes yet found in an insect genome (15). However, as more insect genomes have been sequenced, it has been discovered that *A. mellifera* has an intermediate number of ORs, there is significant variation in OR number among the five ant genomes (16-19), and several solitary insect genomes have among the most ORs found in insects so far (49, 50). Thus, the evidence no longer supports an association between sociality and expansion of the OR repertoire. Furthermore, studies in other organisms have revealed that ORs can function combinatorially and that bioinformatically predicted OR genes may not all produce functional proteins, which together

suggest that the number of ORs in a genome may not scale with the complexity of chemical communication in a species (51).

Due to their functional specificity, ORs are particularly good targets for candidate gene studies because the adaptive significance of OR evolution may be easier to interpret than for genes with broader functions (51). A functional genomics approach was used to identify a novel OR in the *A. mellifera* genome, *AmOr11*, which responds to the main component of the honey bee queen pheromone, (E)-9-oxo-2-decenoic acid (9-ODA) (24). The queen pheromone attracts workers to the queen, partially inhibits worker ovary development, and acts as a sex pheromone, among other functions (24). The specific molecular characteristics of *AmOr11* that are involved the perception of 9-ODA are not yet known, but it appears that it arose early in *Apis* evolution (52 - 54).

Termite queen pheromone. *Neofem2* is the first gene discovered in termites that is involved in signaling queen presence to workers. It was originally identified as being upregulated in female neotenic "replacement" reproductives relative to other colony members, in two species of *Cryptotermes* termites (25). Knocking down *Neofem2* in *Cryptotermes secundus* queens using RNAi caused an increase in aggressive behavior among workers, which is typically only exhibited under queenless conditions (25). Based on sequence similarity, *Neofem2* is most closely related to a β-glycosidase expressed in the salivary glands of the termite *Neotermes koshunensis* (26). β-glycosidases are enzymes that break down polysaccharides; in wood-dwelling termites such as *N. koshunensis* and *C. secundus*, whose diet primarily consists of rotting bark, these enzymes are important for breaking down cellulose (55). It has thus been suggested that *Neofem2* evolved from wood-digesting enzyme to pheromone (26). Supporting this speculation, β-glycosidases exhibit pheromonal activity in other insects, including the production of an egg recognition signal in another termite species (26). The specific molecular changes that have occurred in *Neofem2* as it evolved this new social function remain to be discovered. The story

of *Neofem2* highlights the importance of considering the ecological context of social evolution in a given lineage, as the origin of a social pheromone from a wood-digesting enzyme is almost certainly a phenomenon specific to the wood-dwelling termites.

***Gp-9* in fire ants.** *General Protein-9* (*Gp-9*) alleles are strongly associated with variation in queen number in fire ants (genus *Solenopsis*). In monogynous (single queen) colonies, all females are homozygous for *B*-type alleles and will not tolerate the presence of multiple queens, whereas in polygynous (multiple queens) colonies, some individuals possess *b*-type alleles and do accept multiple queens, but only if those queens also posses the *b*-type allele (27). *Gp-9* has been called a "green beard gene" (28), because workers carrying one allele favor queens that share the same allele. Molecular phylogenetic analysis of *Gp-9* both within and across *Solenopsis* species have revealed that the *b*-like alleles form a monophyletic clade, suggesting that monogyny was the ancestral condition in the genus, and polygyny arose once and has been maintained through multiple speciation events (29, 30).

At the protein sequence level, Gp-9 most closely resembles odorant binding proteins (OBPs), which are expressed in chemosensory sensilla lymph and bind and transport soluble odorants (27). These results have led to the suggestion that *Gp-9* is an OBP that plays a role in pheromonal communication in fire ants (27). However, *Gp-9* is ubiquitously expressed in the hemolymph, suggesting it may be involved in functions that are unrelated to chemosensation (31). In addition, *Gp-9* is found in a genomic region with low recombination rate, and therefore, other linked genes in the region may potentially have more influence on the regulation of queen number (27, 29). *Gp-9* alleles are also associated with variation in several life history traits in *Solenopsis* queens, including body fat and dispersal behavior (30), suggesting that *Gp-9* either acts pleiotropically or with other genes in the region.

Although the function of *Gp-9* is unresolved, molecular evolutionary analyses suggest that this gene is evolving adaptively, implying that *Gp-9* played an important role in fire ant evolution. A signature

of positive selection was detected in the branch leading to the *b*-like allele clade (29) suggesting that this allele had an adaptive benefit when it arose. In addition, all *b*-like alleles share the same amino acid residues at three diagnostic codon positions and two of these positions show evidence of positive selection in *S. invicta*, the species where it has been best studied (30).

Brain development and function

Some of the most striking differences between social and solitary insects are behavioral. Several social insect behaviors appear to be truly novel, such as the dance language in honey bees and slave making in ants (1). Other behaviors exhibited by social insects appear to be modified forms of behaviors performed by solitary insects, for example social foraging, which resembles nest provisioning in solitary insects. It is likely that molecular changes affecting nervous system development and function were important in the evolution of social insect behaviors, but very little is currently known.

Brain evolution in primitively eusocial bees. Our bee molecular evolution study detected a strong signal of rapid evolution in brain-related genes in primitively eusocial, but not in highly eusocial lineages, across two independent origins of each lifestyle (21). Among these rapidly evolving genes were *dunce* and *nejire*, two genes that mediate learning and memory in invertebrates and vertebrates through cAMP/CREB signaling pathways (32).

The detection of molecular changes in brain-related genes exclusively in primitively eusocial bee lineages is perhaps surprising, given that this finding is not what may have been predicted by a prominent hypothesis about the relationship between sociality and brain evolution in vertebrates, the Social Brain Hypothesis (SBH). Originally developed to explain the evolution of the enlarged neocortex in many social vertebrates, the SBH posits that the cognitive demands of social living are a strong selective force in brain evolution (56). Given that highly eusocial bee societies have larger colony sizes, greater

social complexity, and novel behaviors (i.e., the symbolic waggle dance in honey bees) relative to primitively eusocial bees, one might have assumed that the cognitive demands of social living are strongest in highly eusocial species and lead to stronger selection on brain-related genes.

Unique features of insect sociality and the primitively eusocial lifestyle may help explain why selection on brain evolution appears to have been stronger in the primitively eusocial bees. First, unlike in vertebrate social evolution where there has been an emphasis on increased individual cognitive abilities, in insect social evolution there appears to have been an emphasis on increased connectedness among colony members, often accompanied by a reduction of individual behavioral repertoires (57, 58). Therefore, individual cognitive abilities may not be correlated with group size in social insects, as has been found in vertebrates. There are also several distinguishing features of the primitively eusocial bee lifestyle may have placed unique selective pressure on brain evolution in these lineages. Social structure in primitively eusocial bee colonies is typically maintained through fluid and dynamic dominance hierarchies (9, 59), which can be an especially cognitively challenging form of social interaction (59, 60). In addition, a primitively eusocial bee queen is capable of behaving both solitarily, as she does during the colony-founding phase of her lifecycle, and socially, as she does once she has reared her first brood of workers (9).

In both ants and wasps, which each evolved eusociality independently from bees, there are some species in which queens exhibit a similar “solitary-like” phase during colony founding and other species that found colonies in swarms, like highly eusocial bees (1). A comparison of brain-related genes and/or brain structure in ant and wasp species that do or do not establish colonies solitarily may provide clues as to whether this trait is a strong force in social insect brain evolution. One study in paper wasps reported brain region volume differences between swarm and independent-founding species, suggesting that these differences in colony-founding can affect brain evolution (61).

Immunity

Pathogens and parasites are thought to have been a strong a selective force challenging the maintenance of sociality in a variety of organisms, including social insects (62). Crowded living conditions, often with closely related individuals, facilitate pathogen transmission (62). Social insects appear to have responded to this potentially dissolutive selective pressure in three main ways (33). The first way is through "social immunity," which refers to group-level defenses such as hygienic behaviors and the use of collected antimicrobial resins for lining nest cavities (62). The second way is through increasing intracolonial genetic diversity via multiple mating by queens (63) and high rates of genetic recombination (4) to enhance colony-level disease resistance. The third way is through adaptive evolution of immune genes (33).

Molecular evolutionary analyses of immune genes have provided some of the best examples of positive selection acting in social insect genomes. This may be partly due to the fact that immune systems in general are often at the forefront of an ongoing evolutionary arms race with pathogens, so selection pressure on immune-related genes is typically quite strong (64). In addition, many immune-related genes are functionally well-characterized (65), facilitating interpretations of the adaptive significance of sequence changes.

Immune gene evolution in Hymenoptera. When the first social insect genome was sequenced, that of *A. mellifera*, researchers were intrigued by the relatively low number of immune genes found in *A. mellifera* relative to other fully sequenced insect genomes, those of the Diptera, *Drosophila melanogaster* and *Anopheles gambiae* (15). While the main components of canonical immune pathways are conserved, the *A. mellifera* genome contains smaller numbers of gene family members at all points along these pathways (66). It was hypothesized that the loss of immune genes was facilitated by novel forms of social immunity in social insects, resulting in relaxed constraint on immune genes (66).

However, as more insect genomes have been sequenced, it has become apparent that sociality is not necessarily predictive of immune gene number. Rather it seems that Dipterans have unusually large immune gene repertoires, whereas the recently sequenced ant genomes (16-19), the solitary wasps, *Nasonia* (20), and the solitary pea aphid, *Acyrthosiphon pisum* (67) have similar numbers of immune genes as *A. mellifera* (66).

By contrast, molecular evolutionary analysis of individual immune genes in social Hymenoptera has provided evidence that sociality has driven immune gene sequence evolution. One study revealed that some immune genes are evolving more rapidly in species of honey bees, bumble bees and ants relative to *Drosophila* (68). This study also showed that immune genes are evolving more rapidly than nonimmune genes in several honey bee species. Similarly, genes related to innate immunity and humoral immunity were among the fastest evolving (based on branch lengths in phylogenetic trees inferred from protein sequence) in *A. mellifera* in a comparison of over 3,000 genes among *A. mellifera, Nasonia*, and their common ancestor (20). Additionally, evidence for positive selection has been detected in the antimicrobial protein, *defensin*, in a study comparing sequence of 27 ants species (33). This study revealed that the signal and propeptide regions of *defensin*, which are cleaved off to activate the mature peptide, are evolving neutrally, whereas the active region of the peptide is under positive selection, including one amino acid site thought to mediate antimicrobial activity. Our bee molecular evolution study did not detect a strong signal of selection on immune genes, but that was likely because these classes of genes were underrepresented in our dataset (21).

Immune gene evolution in termites. A study of the termite defensin-like gene, *termicin,* in 11 *Nasutitermes* termite species revealed that this gene has duplicated repeatedly during *Nasutitermes* radiation and positive selection has driven a divergence in the molecular charge of the gene copies (34). Insect defensins are known to function by disrupting bacterial plasma membranes and experimental

evidence suggests that molecular charge may be a crucial component of this activity (34). It was hypothesized that there is a selective advantage to having two termicins with different charge properties at specific sites (34). In support of this hypothesis, results from this study suggest that ancestral termicins had relatively high positive charges, and in species in which there has been a gene duplication event, positive selection has driven a decrease in charge for one of the copies. Sequence analysis revealed a strong, positive correlation between the strength of selection (d_N/d_S) and the change in molecular charge along different termicin lineages. Additionally, three amino acid sites that show a signature of positive selection, have substitutions at these sites that contribute to a charge change and they fall on the external surface of the predicted protein structure, suggesting that these sites may interact with a fungal membrane receptor (34).

A different study of 13 *Nasutitermes* termite species also found evidence that gene duplication and positive selection are involved in termite immune gene evolution (36). This study focused on genes encoding Gram-negative bacterial-binding protein 1 and 2 (GNBP1 and GNBP2), which are thought to have duplicated early in termite evolution, and the transcription factor *relish*, which induces production of antimicrobial peptides in *Drosophila*. All three genes show evidence of positive selection, with *relish* showing the strongest signal. Four of the five positively selected sites in *relish* are in a "spacer" region of the protein that is cleaved by the caspase Dredd. This cleavage is thought to activate *relish* by generating a DNA-binding Rel homology domain that translocates to the nucleus and binds to promoters of target genes (69). Analysis of the *Drosophila simulans* ortholog also found positive selection in this spacer region (36). It was hypothesized that microbial pathogens may be targeting this region of *relish* to prevent its activation, sparking an evolutionary arms race as *relish* evolves counter responses to maintain its normal function (36). Another study found evidence of positive selection in *termicin*, but not in GNBP2 in two *Reticulitermes* termite species, a genus distantly related to the *Nasutitermes* genus (35). This study used a population genetics approach to analyze intraspecific polymorphism and

interspecific divergence in coding sequence, and results indicated that *termicin* underwent a selective sweep driven by positive selection for beneficial amino acid changes.

Reproduction

In many insect societies, queens are highly reproductive individuals, whereas workers perform almost no reproduction. Worker sterility is achieved through a variety of morphological, behavioral, and physiological mechanisms in social insects (1). For example, in many social species, workers lack spermatheca for sperm storage. In addition, ovary development is tightly regulated by social cues, and queens and workers typically have grossly over- and underdeveloped ovaries, respectively, relative to solitary insects (1). Sociality also has strong implications for reproductive behavior, particularly for mating frequency, which can affect genetic variation among colony members.

Ovary development in primitively eusocial bees. Our bee molecular evolution study identified some genes involved in ovary development evolving most rapidly in primitively eusocial bees (21). Although both highly and primitively eusocial bee societies have a strong reproductive division of labor, the reproductive differences between queen and worker in primitively eusocial species are less extreme, and ovary development appears to be more sensitive to social cues in primitively eusocial species (1). Perhaps the molecular changes in ovary development-related genes found only in the primitively eusocial lineages underlie some of the unique characteristics of the reproductive biology of this eusocial lifestyle.

Several genes (*tudor, capsuleen, vasa*) evolving rapidly in one or both of the primitively eusocial bee lineages interact together in the PIWI RNA (piRNA) pathway. The piRNA pathway is expressed only in gametic tissue, and is involved in regulating gametic cell division and differentiation (37). Functional PIWI genes have recently been discovered in *A. mellifera* (70), suggesting that the piRNA pathway is

present and functional in bees. These genes are particularly good candidates for further study, as the tissue specificity of the piRNA pathway suggests that selection on these genes is specifically directed at changes related to reproductive processes, in contrast to genes with broader ranges of tissue expression where the functional target of selection is harder to infer. Additional ovary development-related genes unrelated to the piRNA pathway also showed a signature of rapid evolution in these primitively eusocial bees (21).

***csd* and sex determination in honey bees**. More is known about the evolution of *complementary sex determiner* (*csd*) in honey bees than probably any other gene in the social insects. The story of *csd* involves the origin of entirely new genes and pathways, as well as a classic example of balancing selection. Sex determination in honey bees is based on genotype at the *csd* locus; individuals heterozygous at the *csd* locus develop into females, while hemizygous individuals develop into males (38). Sex determination in many Hymenoptera is probably determined through a similar, single-locus system of complementary sex determination (71), but *csd* is the first and only locus that has been discovered thus far. The genomic region containing *csd* was first identified through mapping, and the function of the gene confirmed by RNAi, which showed that genetically female eggs with reduced *csd* expression develop male gonadal tissue (38). Complementary sex determination not only regulates sex determination, but also influences many aspects of social insect biology that are influenced by kinship and degrees of relatedness, including kin selection and the genetic composition of colonies, which is important for division of labor and colony immunity (4).

csd appears to be a honey bee-specific gene, as it has been found in multiple *Apis* species (39), but not outside of the genus (40). The gene likely evolved through the duplication of an adjacent gene, *feminizer* (*fem*). *csd* and *fem* are similar (>70%) in amino acid sequence, and both are serine/arginine-rich proteins (SR-type), a class of proteins involved in RNA splicing (38). Both genes share two major

domains, but *csd* has an additional hypervariable region located between these other domains (40). *fem* has been found in several non-honey bee species and in *Nasonia* wasps, but not in any additional insect species, suggesting that it evolved sometime before the split between the hymenopteran superfamilies Apoidea and Chalcidoidia, ~140 mya, but after the split from *Drosophila*, ~300 mya (40). *fem* shares some functional and sequence similarities to *transformer* (*tra*), a gene involved in sex determination in *Drosophila*, and perhaps evolved from an ancestral form of *tra* common to fly and bee lineages (38, 40). RNAi experiments were used to show that *csd* acts upstream of *fem* in the sex determination pathway. Genetically female embryos treated with *fem* RNAi develop male heads, and RNAi knockdowns of *csd* cause male-specific *fem* splicing, suggesting that *csd* is involved in *fem* splicing (38, 40).

csd has been subject to rigorous population genetic analysis. Because homozygous males do not reproduce, it was predicted that there would be strong negative frequency-dependent selection at the *csd* locus (39). This prediction has been upheld, as at least 15 different *csd* alleles have been found in natural populations around the world in three different *Apis* species (39), and the gene has accumulated 10-13 times more mutations than the rest of the genome (39). Pairwise nonsynonymous differences (πN) between alleles are highest in exons 6 and 7 (39), suggesting that this region is a target of positive selection, and is therefore presumably functionally important. Six fixed amino acid differences between *csd* and *fem* are located in the coiled-coil domain, which are important in protein binding (40). Strong positive selection was detected on the branch right after the split between the two genes, suggesting that positive selection played a role in their diversification (40).

Metabolism and nutrition

Transcriptomic analyses have shown that nutritional and metabolic pathways play an important role in queen-worker caste determination in every eusocial insect lineage thus far studied, and also contribute to worker-worker division of labor in many species (4). Given these fundamental connections

to eusociality, nutritional and metabolic pathways are well-studied in social insects and several molecular evolutionary studies have identified changes associated with their function.

Major Royal Jelly Proteins. The evolution of the Major Royal Jelly Proteins (MRJPs) in honey bees is an excellent example of novel genes playing an integral role in the social biology of a species. In the honey bee *A. mellifera*, the developmental fate of female larvae is determined by the amount of royal jelly they consume (72). Royal jelly is a protein- and lipid-rich substance secreted from the hypopharyngeal glands of brood-feeding "nurse" bees and fed to larvae, which triggers endocrine and epigenetic events that lead to either the development of a worker or a queen (72, 73). The main components of royal jelly are the MRJPs. The *A. mellifera* genome contains 9 MRJPs, encoded by the *mrjp* genes 1-9. These genes are arranged in tandem in the genome, have high sequence similarity (~60%) to one another, and have a conserved intron/exon structure, suggesting that they are a fairly young gene family (41). There is evidence that *mrjp* genes are also present in other *Apis* species (41, 74).

The *mrjp* gene family in *A. mellifera* appears to have evolved via a gene duplication event from a member of the *yellow* gene family. The cluster of *mrjp* genes in the *A. mellifera* genome is flanked by members of the *yellow* gene family, and one of the flanking *yellow* genes, *yellow-e3*, shares the characteristic intron/exon structure of the *mrjp* genes, suggesting that it is their progenitor (41). As the name suggests, members of the *yellow* gene family are typically associated with pigmentation in insects, although they have also been shown to play a role in behavior and sex-specific reproductive maturation in some insects (75).

The use of *mrjp* genes for larval feeding appears to be a derived social trait that is unique to honey bees. Although *mrjp*-like genes have been found in other social and non-social Hymenoptera species, evidence suggests that the *yellow* gene family is prone to duplication, and that the *mrjp*-like genes in non-*Apis* species evolved independently from *Apis* (19, 20). Furthermore, there is no evidence

of a food-related role for any *mrjp*-like or *yellow*-like gene outside of *Apis* (75). As many other social insect species manipulate larval nutrition for the purposes of caste determination without the use of specialized glandular secretions (76), the evolution of the *mrjp* genes in honey bees appear to be associated with the elaboration of eusociality and may have been correlated with or dependent upon other evolutionary changes, such as changes in gland function.

Hexamerins. The work done on the termite hexamerins is another excellent example of linking genetic changes to protein function and social phenotype. In the lower termites, workers may develop into either reproductives or soldiers, depending on a number of social and environmental cues and differentiation into the soldier caste is induced by high juvenile hormone (JH) titers (42). RNAi studies in the termite *Reticulitermes flavipes* have shown that two hexamerin genes, *Hex*-1 and *Hex*-2, are involved in the regulation of this caste determination (43). In many insects, hexamerins act as storage proteins that sequester substances from the diet, and release them when food is scarce or inaccessible, such as during early development (42). It has been hypothesized that *Hex*-1 and *Hex*-2 work together to regulate caste differentiation in termites via direct interactions with JH (43), however elucidating the specific molecular mechanisms involved in JH action is a difficult challenge in insects, in general (77).

Molecular evolutionary studies of *Hex*-1 and *Hex*-2 provide clues as to how these genes may interact with JH. Relative to 100+ known *Hex* genes in other insects, both termite *Hex* genes have distinctive insertions in their coding regions; the unique insertion in *Hex*-1 contains a prenylation motif with a proposed function in JH-binding and the unique insertion in *Hex*-2 shares sequence similarities to the well-characterized blowfly *(Calliphora vicina*) hexamerin receptor (43). Consistent with these predicted functions, follow-up experiments demonstrated that the Hex-1 protein has strong binding affinity for JH, and the Hex-2 protein shows strong membrane affinity, as would be expected for a receptor protein (43).

Hexamerins also exhibit novel social functions in other social insect species, suggesting that they may be particularly prone to social co-option. Evidence in honey bees (78) and *Polistes* wasps (79) suggests that hexamerins may be important in caste determination in these social insect lineages, and in ants, hexamerins appear to be have been important in the evolution of elaborated life history characteristics (80).

The Juvenile Hormone, insulin, and vitellogenin axis. In the highly eusocial honey bee *A. mellifera*, the JH and insulin/insulin-like growth factor-1 (IIS) signaling pathways, and the yolk protein precursor *vitellogenin* (*vg*), interact with one another and function in novel ways that are important in multiple social contexts. JH does not function as a gonadotropin in adult honey bees as it does in most insects, and instead plays a strong role in caste determination and worker division of labor (81). The IIS signaling pathway interacts with JH and is also involved in worker division of labor. Foragers exhibit higher expression of genes in the IIS pathway in the brain relative to nurses and downregulating IIS signaling delays the age-related transition from nursing to foraging (82). This represents a reversal of the traditional positive relationship between high nutrition and IIS signaling as foragers are nutritionally deprived relative to nurses (82). Vg also shows novel social functions in honey bees. It is highly expressed in some workers although they are largely nonreproductive, it may be used by nurses in the synthesis of royal jelly (83) and it functions as an antioxidant that may be involved in promoting longevity in queen bees (84).

The molecular changes underlying these novel functions of JH, IIS, and Vg are unknown, but insights from solitary insects may provide clues as to what these changes may be. The relationship between genetic variation and regulation of JH titers has been particularly well-studied in crickets and butterflies (85), molecular evolution and function of the IIS pathway have been investigated across the

complete genomes of 12 *Drosophila* species (86, 87), and insect vitellogenins and their receptors are well-characterized at the molecular level (88).

Carbohydrate metabolism. Several studies in bees suggest that the evolution of the highly eusocial lifestyle involved molecular changes in genes related to carbohydrate metabolism. Our bee molecular evolution study revealed that genes involved in carbohydrate metabolism are evolving more rapidly in eusocial relative to non-eusocial bee lineages, and are evolving most rapidly in highly eusocial lineages (21). In particular, 15 genes encoding glycolytic enzymes showed evidence of rapid evolution in eusocial lineages, including enzymes that play a key regulatory role (e.g., *phosphofructokinase*) or are involved in glycolytic flux (e.g., *hexokinase*, *pyruvate kinase*) (44). Analysis of protein sequence evolution of genes with queen-biased brain gene expression in *A. mellifera*, found that queen-biased genes that are involved in metabolism, including carbohydrate metabolism, were among the most rapidly evolving (based on branch lengths in phylogenetic trees inferred from protein sequence) relative to orthologs from several solitary insects (89). Comparative analysis of the genome sequences of *A. mellifera, Drosophila melanogaster*, and *Anopheles gambiae* suggest that there may also have been bee-specific changes in gene copy number for carbohydrate metabolizing genes (44). Given that carbohydrate metabolism is such a fundamental, "housekeeping" process, it is not immediately clear why there has been unique selective pressure on these processes in highly eusocial bee lineages. Here, I offer three speculative hypotheses.

First, increases in the flight demands of highly eusocial bees may have placed strong selective pressure on increasing efficiency of glycolytic enzymes, as carbohydrates are the main fuel for flight in bees (90). The individual foraging activity of highly eusocial bee workers appears to be higher than for solitary bees (91), although to the best of our knowledge, no direct comparisons of highly and primitively eusocial bee foraging activity have been performed.

Second, highly eusocial bees are unique in relying exclusively on a diet of modified stored sugars (i.e., honey) for long periods of time. Nest thermoregulation during winter months is completely reliant on honey stores as a fuel source to sustain workers who shiver to produce metabolic heat to maintain optimal hive temperature (92). Perhaps these differences in diet have placed some novel selective pressure on glycolytic enzymes in highly eusocial lineages.

Third, perhaps the greatly extended lifespan of queens in highly eusocial species evolved through changes in metabolism-related genes, including those involved in carbohydrate metabolism. A connection between reduced metabolic rate and increased lifespan has been shown in many species (93). In the honey bee *A. mellifera*, queens exhibit an age-related reduction in IIS signaling (84), which regulates carbohydrate metabolism. If the molecular changes in carbohydrate metabolism genes in highly eusocial bees were due to selection for extended queen lifespan, it can be predicted that similar molecular changes may also be found in independent social insect lineages that also exhibit extended queen lifespans (1).

Prospects and Challenges

Recent work on molecular evolutionary changes in social insects has identified specific genes, molecular pathways and biological processes that appear to have been shaped by natural selection. Some of these changes can be plausibly associated with the origins, maintenance, or elaboration of eusociality, albeit speculatively.

Two insights emerge from this review. First, it appears that there have been unique genetic changes in different social insect lineages, suggesting that the multiple independent occurrences of eusociality have involved multiple molecular routes. These differences may reflect the ecological or other constraints in which eusociality evolved in each group. For example, the evolution of a queen

pheromone in termites from a wood-digesting enzyme seems fitting, given that many termite societies live in rotting wood (26).

Second, genetic changes also have occurred in similar biological functions across diverse species of social insects. This supports the concept of a genetic toolkit for eusociality (94). This concept is reasonable because despite the striking diversity among social insect species, they all have converged on a similar suite of traits, which are the defining characteristics of eusociality (9). Previous research suggesting components of a genetic toolkit for eusociality has focused on genes and molecular pathways that are associated both with solitary and related social behaviors in insects, for example, the *foraging* gene, which is involved in feeding behavior in *Drosophila* and a variety of other solitary organisms, and social foraging behavior in honey bees and ants (94). Transcriptomic studies have also identified shared sets of genes whose expression patterns are associated with division of labor in independent social insect lineages (95).

The molecular evolutionary studies I reviewed identify biological processes and specific genes that may be excellent systems in which to further investigate the concept of a genetic toolkit for eusociality. Among the most promising are: 1) *Hexamerins*. As discussed above, hexamerins have been shown to be involved in queen physiology and other social traits in a variety of social insects and the work on *Hex-1* and *Hex-2* in termites demonstrates how hexamerin sequence evolution can be studied and linked to social traits. 2) *Gland development genes.* The rapidly evolving gland development genes identified in our bee molecular evolution study (21) are also good candidates for further study because the gene functions are relatively well-characterized and gland diversification is a universal phenomenon in social insect evolution. 3) *Brain-related genes.* The rapidly evolving brain-related genes identified in primitively eusocial lineages in our bee molecular evolution study (21) are prime candidates for further study in primitively eusocial bees, as well as ant and wasp species who share the primitively eusocial bee lifestyle feature of solitary nest-founding.

The molecular changes and biological processes highlighted in this review are currently the most well-studied in social insects. There are almost certainly other equally important types of molecular changes and biological processes associated with social insect evolution that have not yet been discovered, perhaps because of the limited range of taxa subjected to these type of analyses thus far. This gap in our knowledge is largely due to a lack of genomic resources, especially for closely related, social and non-social species. For example, some types of genetic changes, such as chromosomal rearrangements and patterns of DNA methylation, are not possible to study with only fragments of the genome. In addition, the identification of truly novel genes is limited by the small sample size of available genomes and less well developed forward genetic analyses in social insects relative to model genetic organisms. As these limitations are overcome, it should be possible to search more broadly for different types of genetic changes associated with the evolution of eusocial traits. These analyses can be guided by several theoretical models that have been proposed to predict the types of genetic changes that are most important in social evolution (96 – 98).

Whole genome scans for molecular signatures of adaptive evolution specific to social insects will be particularly useful for generating new hypotheses and implicating new biological processes in social insect evolution. Candidate gene approaches across a broad sample of social and non-social insects will allow for greater accuracy in reconstructing the phylogenetic history of molecular changes and testing their associations with social evolution. Once specific sequence changes are identified, functional analyses are necessary to determine their effect on protein-, organismal-, and group-level phenotypes, and the adaptive significance of the phenotype change (99).

This leads me to raise one important caveat for most molecular evolutionary studies in the social insects: the lack of species-specific information about gene function. As is often the case in this paper, gene function is typically inferred solely from orthology to the fruit fly *Drosophila melanogaster*, which shared a common ancestor with eusocial insect lineages over 300 million years ago (15). Although

gene function for molecular processes is generally highly conserved over evolutionary time, when interpreting findings, it is important to consider the possibility that a particular gene has evolved a novel function. Furthermore, many genes have multiple functions, and thus the target of selection can be difficult to infer solely from identifying molecular evolutionary changes. Experimental approaches to determining gene function in social insects, via RNAi and transgenesis, will strengthen the interpretation of molecular evolutionary findings. Additional challenges arise in determining the adaptive or ecological significance of molecular changes, even when their functional significance is understood (100).

Despite these challenges, molecular evolutionary analysis of social insect societies holds promise for testing venerable theories of social evolution using genomic data. Multiple evolutionary scenarios have been proposed as potential routes to group living in insects. These include the composition of incipient social groups such as associations between mothers and offspring (the "subsocial" route) or between related or unrelated individuals of the same generation ("semisocial" route) (9); mechanisms through which altruism is achieved, such as kin selection (2), parental manipulation of offspring, or voluntary helpers at the nest (101); and necessary preadaptations for social living, such as a monogamous mating system (102) or progressive provisioning of offspring (3). Wedding this rich theory with genome-scale molecular evolutionary analysis and functional experimentation holds the promise of finally answering the compelling question of how eusociality evolved in insects.

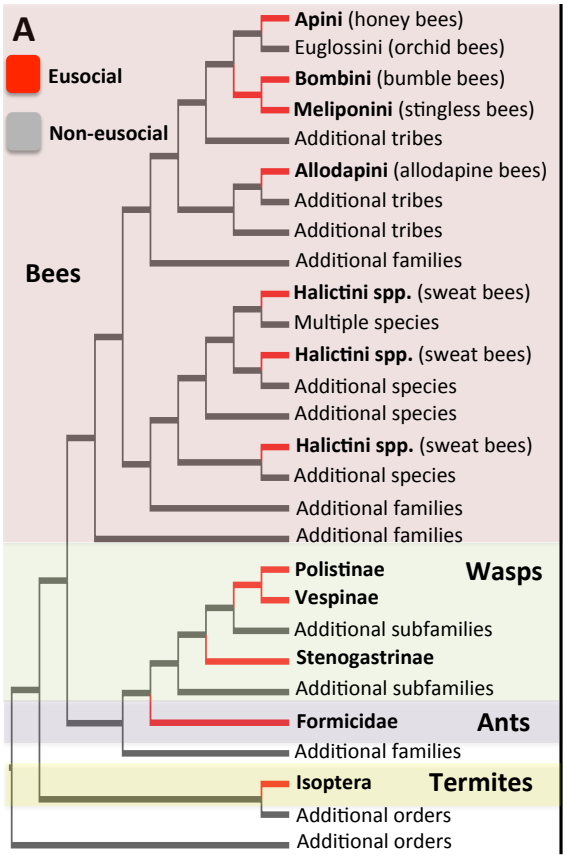
ACKNOWLEDGMENTS

I thank members of the G.E.R. laboratory for comments that improved this manuscript. The bee molecular evolution study was supported by 454 Life Sciences (Roche Diagnostics Corporation) via the Roche 1GB contest, National Science Foundation Grant DEB07-43154 (to G.E.R., Primary Investigator), National Science Foundation Predoctoral Fellowship NSF DGE 07-15088 FLW (to B.J.F.), and National

Institutes of Health–University of Illinois Sensory Neuroscience Training Grant PHS2T32DC006612 (to S.H.W. and A. S. Feng, Primary Investigator).

FIGURES AND TABLES

Figure 1. Cladogram showing the origins of eusociality in insects. Topology and reconstruction of evolutions of eusociality are based on multiple studies (5–8).



B

Bees	Variable social complexity (highly and primitively eusocial species) No intra-worker morphological castes Some losses of eusociality (*e.g.,* subgenus *Psithyrus*)	Live in cavities, in the open, or underground Diet of pollen and nectar
Wasps	All primitively eusocial No intra-worker morphological castes	Live in nests constructed from pulp, other collected materials Herbivorous (diet mostly insects and nectar)
Ants	Incredibly diverse lifestyles (*e.g.,* slavemakers) Intra-worker castes (*e.g.,* soldiers) Some queenless lineages	Live underground or in cavities Herbivorous; some lineages cultivate fungus
Termites	Colonies have both kings and queens Intra-worker castes (*e.g.,* soldiers) Secondary reproductives (neotenics) Only diploid eusocial insect	Live in rotting wood, underground Detritivores; some lineages cultivate fungus

Table 1. Genes implicated in the origin or maintenance of insect society by molecular evolutionary research. Although many genes in this table are presumably involved in multiple biological processes, they are classified in one of five processes with known links to insect sociality: chemical signaling; brain development and function; immunity; reproduction; and metabolism and nutrition.

Gene	Function	Evidence	Type of change*
Chemical signaling			
decapentaplegic	Gland development (22, 23).	Rapid evolution in eusocial bees (21)	1
thickveins	Gland development (22, 23).	Rapid evolution in eusocial bees (21)	1
PDGF- and VEGF-related factor 1	Gland development (22, 23).	Rapid evolution in eusocial bees (21)	1
AmOr11	Odorant receptor (24)	Responds to main component of queen honey bee pheromone, 9-ODA (24)	2
Neofem 2	β-glycosidase-like (25, 26)	Involved in signaling queen termite presence (25, 26)	3
GP-9	Putative odorant binding protein (27-31)	Allelic variation associated with fire ant queen number (27-31)	1, 2
Brain development and function			
dunce	cAMP/CREB signaling pathways (32).	Rapid evolution in primitively eusocial bees (21)	1
nejire	CREB binding protein (32).	Rapid evolution in primitively eusocial bees (21)	1
Immunity			
defensin	Antimicrobial protein (33)	Positive selection in ants (33)	1
termicin	Antimicrobial protein (34, 35)	Gene duplication, positive selection in termites (34, 35)	1, 2
Gram-negative bacterial-binding protein 1 and *2*	Pattern recognition receptors (36)	Gene duplication, positive selection in termites (36)	1, 2
relish	Transcription factor, induces production of antimicrobial peptides (36)	Positive selection in termites (36)	1
Reproduction			
tudor	PIWI RNA pathway (37)	Rapid evolution in primitively eusocial bees (21)	1
capsuleen	PIWI RNA pathway (37)	Rapid evolution in primitively eusocial bees (21)	1

Table 1 (continued)

vasa	PIWI RNA pathway (37)	Rapid evolution in primitively eusocial bees (21)	1
complementary sex determiner	Sex determination (38-40)	Gene duplication, positive selection in honey bees (38-40)	1, 2
Metabolism and nutrition			
Major Royal Jelly Proteins	Main components of royal jelly (41)	Gene family expansion, novel feeding-related functions in honey bees (41)	2
Hex-1 and *Hex-2*	Storage proteins (42, 43)	Unique insertions in termites (42, 43)	1
phosphofructokinase	Key regulator of glycolysis (44)	Rapid evolution in eusocial bees (21)	1
hexokinase	Regulator of glycolyitic flux (44)	Rapid evolution in eusocial bees (21)	1
pyruvate kinase	Regulator of glycolyitic flux (44)	Rapid evolution in eusocial bees (21)	1

* 1 = protein coding sequence change; 2 = novel gene; 3 = change unknown

LITERATURE CITED

1. Wilson EO (1971) *The Insect Societies* (Belknap Press, Cambridge, MS).
2. Strassmann JE, Queller DC (2007) Insect societies as divided organisms: the complexities of purpose and cross-purpose. *Proc Natl Acad Sci USA* 104:8619-8626.
3. Nowak MA, Tarnita CE, Wilson EO (2010) The evolution of eusociality. *Nature* 466:1057-1062.
4. Smith CR, Toth AL, Suarez AV, Robinson GE (2008) Genetic and genomic analyses of the division of labour in insect societies. *Nat Rev Genet* 9:735-748.
5. Cameron SA, Mardulyn P (2001) Multiple molecular data sets suggest independent origins of highly eusocial behavior in bees (Hymenoptera:Apinae). *Syst Biol* 50:194-214.
6. Cardinal S, Straka J, Danforth BN (2010) Comprehensive phylogeny of apid bees reveals the evolutionary origins and antiquity of cleptoparasitism. *Proc Natl Acad Sci USA* 107:16207-16211.
7. Hines HM, Hunt JH, O'Connor TK, Gillespie JJ, Cameron SA (2007) Multigene phylogeny reveals eusociality evolved twice in vespid wasps. *Proc Natl Acad Sci USA* 104: 3295-3299.
8. Brady SG, Sipes S, Pearson A, Danforth BN (2006) Recent and simultaneous origins of eusociality in halictid bees *Proc Roy Soc Lond B* 273:1643-1649.
9. Michener CD (1974) *The Social Behavior of Bees* (Belknap Press, Harvard).
10. Clark AG (2006) Genomics of the evolutionary process. *Trends Ecol Evol* 21:316-321.
11. Li WH (1997) *Molecular Evolution* (Sinauer Associates).
12. *Drosophila* 12 Genomes Consortium (2007) Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature* 450:203-218.
13. Clark AG, et al. (2003) Inferring nonneutral evolution from human-chimp-mouse orthologous gene trios. *Science* 302:1960-1963.
14. Pollard KS, et al. (2006) An RNA gene expressed during cortical development evolved rapidly in humans. *Nature* 443:167-172.

15. HGBSC (2006) Insights into social insects from the genome of the honeybee *Apis mellifera*. *Nature* 443:931-949.
16. Bonasio R, et al. (2010) Genomic comparison of the ants *Camponotus floridanus* and *Harpegnathos saltator*. *Science* 329:1068-1071.
17. Wurm Y, et al. (2011) The genome of the fire ant *Solenopsis invicta*. *Proc Natl Acad Sci USA.*
18. Smith CR, et al. (2011) Draft genome of the red harvester ant *Pogonomyrmex barbatus. Proc Natl Acad Sci USA.*
19. Smith CD, et al. (2011) Draft genome of the globally widespread and invasive Argentine ant (*Linepithema humile*). *Proc Natl Acad Sci USA.*
20. Werren JH, et al. (2010) Functional and evolutionary insights from the genomes of three parasitoid *Nasonia* species. *Science* 327:343-348.
21. Woodard SH, et al. (2011) Genes involved in convergent evolution of eusociality in bees. *Proc Natl Acad Sci USA* (in press).
22. Bradley PL, Myat MM, Comeaux CA, Andrew DJ (2003) Posterior migration of the salivary gland requires an intact visceral mesoderm and integrin function. *Dev Biol* 257:249-262.
23. Harris KE, Schnittke N, Beckendorf SK (2007) Two ligands signal through the *Drosophila* PDGF/VEGF receptor to ensure proper salivary gland positioning. *Mech Develop* 124:441-448.
24. Wanner KW, et al. (2007) A honey bee odorant receptor for the queen substance 9-oxo-2-decenoic acid. *Proc Natl Acad Sci USA* 104:14383-14388.
25. Weil T, Korb J, Rehli M (2009) Comparison of queen-specific gene expression in related lower termite species. *Mol Biol Evol* 26:1841-1850.
26. Korb J, Weil T, Hoffmann K, Foster KR, Rehli M (2009) A gene necessary for reproductive suppression in termites. *Science* 324:758-758.

27. Gotzek D, Ross K (2009) Current status of a model system: the gene *Gp-9* and its association with social organization in fire ants. *Plos One* 4:e7713.

28. Keller L, Ross KG (1998) Selfish genes: a green beard in the red fire ant. *Nature* 394:573-575.

29. Krieger MJB, Ross K (2005) Molecular evolutionary analyses of the odorant-binding protein gene *Gp-9* in fire ants and other *Solenopsis* species. *Mol Biol Evol* 22:2090-2103.

30. Gotzek D, Shoemaker DD, Ross K (2007) Molecular variation at a candidate gene implicated in the regulation of fire ant social behavior. *Plos One* 2:e1088.

31. Leal WS, Ishida Y (2008) *GP-9*s are ubiquitous proteins unlikely involved in olfactory mediation of social organization in the red imported fire ant, *Solenopsis invicta*. *Plos One* 3:e3762.

32. Silva AJ, Kogan JH, Frankland PW, Kida S (1998) CREB and memory. *Ann Rev Neuro* 21:127-148.

33. Viljakainen L, Pamilo P (2008) Selection on an antimicrobial peptide *defensin* in ants. *J Mol Evol* 67:643-652.

34. Bulmer M, Crozier R (2004) Duplication and diversifying selection among termite antifungal peptides. *J Mol Evol* 21:2256-2264.

35. Bulmer MS, Lay F, Hamilton C (2010) Adaptive evolution in subterranean termite antifungal peptides. *Ins Mol Bio* 19:669-674.

36. Bulmer M, Crozier R (2006) Variation in positive selection in termite GNBPs and Relish. *J. Mol. Evol.* 23:317-326.

37. Siomi MC, Mannen T, Siomi H (2010) How does the royal family of *tudor* rule the PIWI-interacting RNA pathway? *Genes Dev* 24:636-466.

38. Gempe T, et al. (2009) Sex determination in honeybees: two separate mechanisms induce and maintain the female pathway. *Plos Biol* 7:e1000222.

39. Hasselmann M, et al. (2008) Evidence for convergent nucleotide evolution and high allelic turnover rates at the *complementary sex determiner* gene of Western and Asian honeybees. *Mol Biol Evol* 25:696-708.

40. Hasselmann M, et al (2008) Evidence for the evolutionary nascence of a novel sex determination pathway in honeybees. *Nature* 454:519-523.

41. Drapeau MD, Albert S, Kucharski R, Prusko C, Maleszka R (2006) Evolution of the Yellow/Major Royal Jelly Protein family and the emergence of social behavior in honey bees. *Genome Res* 16:1385-1394.

42. Zhou X, Tarver MR, Scharf ME (2007) Hexamerin-based regulation of juvenile hormone-dependent gene expression underlies phenotypic plasticity in a social insect. *Development* 134:601-610.

43. Zhou X, Tarver MR, Bennett GW, Oi FM, Scharf ME (2006) Two hexamerin genes from the termite *Reticulitermes flavipes*: sequence, expression, and proposed functions in caste regulation. *Gene* 376:47-58.

44. Kunieda T, et al. (2006) Carbohydrate metabolism genes and pathways in insects: insights from the honey bee genome. *Insect Mol Biol* 15:563-576.

45. Le Conte Y, Hefetz A (2008) Primer pheromones in the social Hymenoptera. *Ann Rev Entomol* 53:523-542.

46. Enard W, et al. (2002) Molecular evolution of *FOXP2*, a gene involved in speech and language *Nature* 418:869-872.

47. Wagner GP, Pavlicev M, Cheverud JM (2007) The road to modularity. *Nat Rev Genet* 8:921-31.

48. Robertson HM, Wanner KW (2006) The chemoreceptor superfamily in the honey bee, *Apis mellifera*: expansion of the odorant, but not gustatory, receptor family. *Genome Res* 16:1395-1403.

49. Engstonia P, et al. (2008) The red flour beetle's large nose: an expanded odorant receptor gene family in *Tribolium castaneum*. *Insect Biochem Mol Biol* 38:387-397.

50. Robertson HM, Gadau J, Wanner KW (2010) The insect chemoreceptor superfamily of the parasitoid jewel wasp *Nasonia vitripennis. Insect Mol Biol* 19:121-136.
51. Nei M, Niimura Y, Nozawa M (2008) The evolution of animal chemosensory receptor gene repertoires: roles of chance and necessity. *Nat Rev Genet* 9:951-963.
52. Plettner E, et al. (1997) Species- and caste-determined mandibular gland signals in honeybees *(Apis). J Chem Ecol* 23:363-377.
53. Urbanova K, Cahlikova L, Hovorka O, Ptacek V, Valterova I (2008) Age-dependent changes in the chemistry of exocrine glands of *Bombus terrestris* queens. *J Chem Ecol* 34:458-466.
54. Cruz-Lopez L, et al. (2005) Mandibular gland secretion of *Melipona beecheii*: chemistry and behavior. *J Chem Ecol* 31:1621-1632.
55. Tokuda G, Saito H, Watanabe H (2002) A digestive β-glucosidase from the salivary glands of the termite, *Neotermes koshunensis* (Shiraki): distribution, characterization and isolation of its precursor cDNA by 5- and 3-RACE amplifications with degenerate primers. *Insect Biochem Mol Biol* 32:1681-1689.
56. Dunbar RIM, Shultz S (2007) Evolution in the social brain. *Science* 317:1344-1347.
57. Oster GF, Wilson EO (1979) *Caste and Ecology in the Social Insects* (Princeton University Press).
58. Gronenberg W, Riveros AJ (2009) In *Organization of Insect Societies: From Genome to Sociocomplexity*, eds Gadau J, Fewell J (Harvard University Press), pp 377-401.
59. O'Donnell S, Donlan N, Jones T (2007) Developmental and dominance-associated differences in mushroom body structure in the paper wasp *Mischocyttarus mastigophorus*. *Dev Neurobiol* 67:39-46.
60. Salvador A, Costa R (2009) Coping with competition: neuroendocrine responses and cognitive variables. *Neurosci Biobehav Rev* 33:160-170.

61. Molina Y, Harris RM, O'Donnell S (2009) Brain organization mirrors caste differences, colony founding and nest architecture in paper wasps (Hymenoptera: Vespidae). *Proc Roy Soc Lond B* 276:3345-3351.

62. Wilson-Rich N, Spivak M, Fefferman NH, Starks PT (2009) Genetic, individual, and group facilitation of disease resistance in insect societies. *Ann Rev Entomol* 54:405-423.

63. Tarpy D, Seeley TD (2006) Lower disease infections in honeybee (*Apis mellifera*) colonies headed by polyandrous vs monandrous queens. *Naturwissen* 93:195-9.

64. Lazzaro BP (2008) Natural selection on the *Drosophila* antimicrobial immune system. *Curr Opin Microbiol* 11:284-289.

65. Hoffmann JA (2003) The immune response of *Drosophila*. *Nature* 426:33-38.

66. Evans JD, et al. (2006) Immune pathways and defense mechanisms in honey bees *Apis mellifera*. *Insect Mol Biol* 15:645-656.

67. The International Aphid Genomics Consortium (2010) Genome sequence of the pea aphid *Acyrthosiphon pisum*. *PLoS Biol* 8:e1000313.

68. Viljakainen L, et al. (2009) Rapid evolution of immune proteins in social insects. *Mol Biol Evol* 26:1791-801.

69. Stoven S, et al. (2003) Caspase-mediated processing of the *Drosophila* NF-κB factor Relish. *Proc Natl Acad Sci USA* 100:5991-5996.

70. Liao Z, Jia Q, Li F, Han Z (2010) Identification of two piwi genes and their expression profile in honeybee, *Apis mellifera*. *Arch Insect Biochem* 74:91-102.

71. Cook JM (1993) Sex determination in the Hymenoptera: a review of models and evidence. *Heredity* 71:421-435.

72. Weaver N (1955) Rearing of honeybee larvae on royal jelly in the laboratory. *Science* 121:509-510.

73. Lyko F, et al. (2010) The honey bee epigenomes: differential methylation of brain DNA in queens and workers. *PLoS Biol* 8:e1000506.

74. Yu F, Mao F, Jianke L (2010) Royal jelly proteome comparison between *A. mellifera* ligustica and *A. cerana* cerana. *J Proteome Res* 9:2207-2215.

75. Ferguson L, Green J, Surridge A, Jiggins CD (2010) Evolution of the insect yellow gene family. *Mol Biol Evol* 28:257-272.

76. Webster T, Peng Y (1988) The evolution of food-producing glands in eusocial bees (Apoidea, Hymenoptera). *J Evol Biol* 2:165-176.

77. Riddiford LM (2008) Juvenile hormone action: a 2007 perspective. *J Insect Physiol* 54:895-901.

78. Martins JR, Nunes FMF, Cristino AS, Simoes ZLP, Bitondi MMG (2010) The four hexamerin genes in the honey bee: structure, molecular evolution and function deduced from expression patterns in queens, workers and drones. *BMC Mol Biol* 11:23.

79. Hunt JH, et al. (2010) Differential gene expression and protein abundance evince ontogenetic bias toward castes in a primitively eusocial wasp. *Plos One* 5: e10674.

80. Wheeler DE, Buck NA (1995) Storage proteins in ants during development and colony founding. *J Insect Physiol* 41:885-894.

81. Robinson GE, Vargo EL (1997) Juvenile hormone in adult eusocial Hymenoptera: gonadotropin and behavioral pacemaker. *Arch Insect Biochem Physiol* 35:559-583.

82. Ament SA, Corona M, Pollock HS, Robinson GE (2008) Insulin signaling is involved in the regulation of worker division of labor in honey bee colonies. *Proc Natl Acad Sci USA* 105:4226-4231.

83. Amdam GV, Norberg K, Hagen A, Omholt S (2003) Social exploitation of vitellogenin. *Proc Natl Acad Sci USA* 100:1799-1802.

84. Corona M, et al. (2007) Vitellogenin, juvenile hormone, insulin signaling, and queen honey bee longevity. *Proc Natl Acad Sci USA* 104:7128–7133.

85. Zera AJ, Harshman LG, Williams TD (2007) Evolutionary endocrinology: the developing synthesis between endocrinology and evolutionary genetics. *Annu Rev Ecol Evol Syst* 38:793-817.
86. Grönke S, Clarke DF, Broughton S, Andrews TD, Partridge L (2010) Molecular evolution and functional characterization of *Drosophila* insulin-like peptides. *PLoS Genet* 6:e1000857.
87. Alvarez-Ponce D, Aguade M, Rozas J (2009) Network-level molecular evolutionary analysis of the insulin/TOR signal transduction pathway across 12 *Drosophila* genomes. *Genome Res* 19:234-242.
88. Sappington TW, Raikhel AS (1998) Molecular characteristics of insect vitellogenins and vitellogenin receptors. *Insect Biochem Mol Biol* 28:277-300.
89. Hunt BG, et al. (2010) Sociality is linked to rates of protein evolution in a highly social insect. *Mol Biol Evol* 27:497-500.
90. Suarez RK, et al. (2005) Energy metabolism in orchid bee flight muscles: carbohydrate fuels all. *J Exp Biol* 208:3573-3579.
91. Roubik DW (1992) *Ecology and Natural History of Tropical Bees* (Cambridge University Press).
92. Southwick EE, Heldmaier G (1987) Temperature control in honey bee colonies. *BioScience* 37:395-399.
93. Finkel T, Holbrook NJ (2000) Oxidants, oxidative stress and the biology of ageing. *Nature* 408:239-247.
94. Toth AL, Robinson GE (2009) Evo-devo and the evolution of social behavior: brain gene expression analyses in social insects. *Cold Spring Harb Symp Quant Biol* 74:419-426.
95. Toth AL, et al. (2010) Brain transcriptomic analysis in paper wasps identifies genes associated with behaviour across social insect lineages. *Proc Roy Soc Lond B* 277:2139-2148.
96. Nonacs P, Kapheim KM (2008) Social heterosis and the maintenance of genetic diversity at the genome level. *J Evol Biol* 20:2253-2265.

97. Linksvayer TA, Wade MJ (2009) Genes with social effects are expected to harbor more sequence variation within and between species. *Evolution* 63:1685-96.

98. Johnson BR, Linksvayer TA (2010) Deconstructing the superorganism: social physiology, groundplans, and sociogenomics. *Q Rev Biol* 85:57-79.

99. Dean AM, Thornton JW (2007) Mechanistic approaches to the study of evolution: the functional synthesis. *Nat Rev Genet* 8:675-688.

100. Feder ME, Mitchell-Olds T (2008) Evolutionary and ecological functional genomics. *Nat Rev Genet* 4:651-657.

101. Charnov EL (1978) Evolution of eusocial behavior: offspring choice or parental parasitism? *J Theor Biol* 75:451-465.

102. Hughes WOH, Oldroyd BP, Beekman M, Ratnieks FLW (2008) Ancestral monogamy shows kin selection is key to the evolution of eusociality. *Science* 320:1213-1216

CHAPTER 3: GENES INVOLVED IN CONVERGENT EVOLUTION OF EUSOCIALITY IN BEES [1]

Abstract

Eusociality has arisen independently at least eleven times in insects. Despite this convergence, there are striking differences among eusocial lifestyles, ranging from species living in small colonies with overt conflict over reproduction, to species in which colonies contain hundreds of thousands of highly specialized sterile workers produced by one or a few queens. Although the evolution of eusociality has been intensively studied, the genetic changes involved in the evolution of eusociality are relatively unknown. I examined patterns of molecular evolution across three independent origins of eusociality by sequencing transcriptomes of nine socially diverse bee species, and combining these data with genome sequence from the honey bee *Apis mellifera*, to generate orthologous sequence alignments for 3647 genes. I found a shared set of 212 genes with a molecular signature of accelerated evolution across all eusocial lineages studied, as well as unique sets of 173 and 218 genes with a signature of accelerated evolution specific to either highly or primitively eusocial lineages, respectively. These results demonstrate that convergent evolution can involve a mosaic pattern of molecular changes in both shared and lineage-specific sets of genes. Genes involved in signal transduction, gland development, and carbohydrate metabolism are among the most prominent rapidly evolving genes in eusocial lineages. These findings provide a starting point for linking specific genetic changes to the evolution of eusociality.

---

[1] Chapter 3 was published in *Proceedings of the National Academy of Sciences* in 2011: Woodard SH, Fischman BJ, Venkat A, Hudson ME, Varala K, Cameron SA, Clark AG, Robinson GE (2011) Genes involved in convergent evolution of eusociality in bees. *Proc. Natl. Acad. Sci. USA* 108: 7472-7477. "Supplement" in thesis refers to supplemental data for publication.

[*] These authors contributed equally to this work

Author contributions: S.H.W. and B.J.F. designed the study, edited sequences, performed evolutionary and GO analyses, analyzed results, and wrote the paper; A.V., M.E.H. & K.V. assembled ESTs, identified orthologs, and performed additional bioinformatic analyses; S.A.C. performed molecular phylogenetic analyses and wrote the corresponding sections; A.G.C. designed the study and analyzed results; and

G.E.R. designed the study, analyzed results, and wrote the paper. All authors discussed the results and commented on the manuscript.

Introduction

The evolution of eusociality, the phenomenon in which female offspring forgo personal reproduction to care cooperatively for their siblings, is one of the major transitions of life on Earth (1). This evolutionary transition has occurred multiple times, but only in a small number of lineages, primarily in the insects (11 or more times) (2). The evolution of eusociality has long fascinated biologists because it requires that the balance between cooperation and conflict shift in favor of cooperation, in spite of strong selective pressure for individual reproductive success (3).

Despite a rich history of theoretical work on the evolution of eusociality (4, 5), relatively little is known about the molecular changes associated with eusocial evolution (6). These molecular changes have the potential to inform us about the evolutionary processes involved in the evolution of eusociality, such as types and levels of selection (7). Some insights have been gained about molecular mechanisms underlying eusociality in individual eusocial lineages (6), but a broad comparative framework for exploring common principles of the molecular basis of eusocial evolution is lacking. One major unresolved question is whether independent evolutionary trajectories of eusociality involved similar or different genetic changes. I explored the genetic basis of eusocial evolution in bees, an ideal group for comparative studies of social evolution. There is a wide diversity of social lifestyles within this group, from solitary to intermediately social to elaborate eusociality (8). Additionally, eusociality has been gained independently at least six times (9-12) in the bees, more than in any other group. These features make it possible to compare multiple, independent origins of different social lifestyles among relatively closely related species. Furthermore, the extensive knowledge of bee natural history (8, 13, 14) provides a valuable framework for developing hypotheses about the adaptive significance of genetic changes detected in eusocial bee lineages.

To study patterns of molecular evolution associated with eusociality in bees, I generated ~1 Gbp of Expressed Sequence Tags (ESTs) from a set of nine bee species. This set of species reflects the remarkable social diversity in bees by including eusocial and non-eusocial species, three origins of eusociality (9,10), and two different forms of eusocial lifestyle, "highly eusocial" and "primitively eusocial" (8) (Fig. 2A). I combined the ESTs with genome sequence from the highly eusocial honey bee *Apis mellifera* (15) and created manually curated, 10-species, partial gene sequence alignments. I searched among the alignments for genes with accelerated rates of amino acid substitution in eusocial relative to non-eusocial lineages. Accelerated rates of protein evolution can reflect a molecular signature of positive natural selection (16), and shared patterns of acceleration among lineages can suggest an association between genetic changes and the evolution of shared traits.

Methods

Bee collection and sequencing. Bees used for sequencing were free-flying or collected from nests. They were placed directly into liquid nitrogen for RNA preservation. Different ages, behavioral groups, and castes (when applicable) were used to maximize transcript diversity. RNA was extracted from brains and abdomens of 50+ females per species. Pooled mRNA (90% brain, 10% abdomen) was sequenced by 454 Life Science/Roche on the GS-FLX platform. Most transcripts in the genome are expressed in the brain (17); abdomen tissue was added to enhance transcript discovery for reproduction-related processes. Additional information about collections, RNA extractions, and sequencing is in Supplement.

EST and alignment assembly. EST reads were assembled using Phrap to generate species-specific, non-redundant contigs and singletons. *A. mellifera* gene models were obtained from BeeBase (Official Honey Bee Gene Set http://genomes.arc.georgetown.edu/drupal/beebase/). For each species, the assembled ESTs were matched to the *A. mellifera* gene models. Orthology was determined using the reciprocal best

BLAST hit. Gapped ortholog-reference-guided transcript assemblies (GOTAs) were created by concatenating the top reciprocal hits and trimming the overlaps. Multiple sequence alignments were then created using MAFFT software (18). All alignments were manually inspected in Geneious (19), and ambiguous regions and gaps covering whole codons were deleted. Additional information about ortholog assignment and editing is in Supplement.

Phylogeny. Nucleotide sequences for 3647 protein-coding EST gene fragments were aligned (18), edited manually (19), and modified to include fragments containing no gaps for any of the 10 taxa. Gene fragments of length > 100 bp were concatenated and the resulting inframe nucleotide alignment (n = 717 gene fragments; 69461 bp total) was analyzed with Bayesian inference in MrBayes (v3.1.2 MPI (parallel) version for unix clusters) (20) under the substitution model GTR+I+G; amino acid translation analyses were run using the JTT fixed rate model. Figure 2 is the consensus of the Bayesian posterior distribution of phylogenetic trees from analysis of 3rd codon positions. The consensus trees based on all, 1st, and 2nd position nucleotide sites, and amino acid sequence are reported in Supplement.

Evolutionary tests. The program *codeml* in the PAML package (21) was used to fit the alignment data to branch models of codon substitution by maximum likelihood to identify differences in ω within the tree. For each test, the likelihoods of two models of evolution (neutral and non-neutral) were compared using a likelihood ratio test (LRT). Any genes with one or more branches with $dS > 2$ ($n = 9$) were considered to be saturated, and were excluded from further analyses. To correct for multiple tests, an FDR correction on nominal P values obtained from the LRTs was performed.

Gene Ontology enrichment analysis. For functional analyses, a pre-existing list of *A. mellifera-D. melanogaster* orthologs was used (15). Orthologous fly sequences with annotation information were

available for most (*n* = 3451) genes in the dataset. GO analyses were performed using the functional annotation tool on DAVID (22). Additional information about GO analysis is in Supplement.

Results

Characterization of alignments. The alignments corresponded to ~33% of the genes (*n* = 3647; 3638 after removal of alignments showing evidence of saturation) in the *A. mellifera* Official Gene Set. To improve the utility of this genomic resource for evolutionary analysis, I used stringent criteria for assessing orthology to minimize misclassification of paralogous sequences within the alignments (see Supplement). I also looked for functional biases in the set of genes represented by our alignments by performing Gene Ontology enrichment analysis. I identified biological processes that were over- and under-represented in the set of genes relative to all genes in the *A. mellifera* Official Gene Set (data not shown).

Phylogenetic tree inference from EST data. Bayesian inference was used to estimate the phylogenetic relationships among bee species from our set of 3638 alignments (see Supplement). The phylogenetic tree inferred from 3rd nucleotide positions was identical in structure to trees inferred in previously published studies that included greater taxonomic sampling (7, 8) (see Supplement). A single, different topology was obtained by inferring phylogeny from the other nucleotide positions and from amino acid sequences (see Supplement). I therefore performed all of our molecular evolutionary analyses using both tree topologies. Overall, tree topology had little effect on the results of our molecular evolutionary analyses (see Supplement), and the results reported here use the topology in Fig. 2.

Heterogeneous patterns of molecular evolution among bee lineages. I searched among the alignments for genes with accelerated rates of amino acid substitution in eusocial relative to non-eusocial lineages. I

performed two tests (Figure 2B, C) that used Likelihood Ratio Tests (LRTs) to compare models of neutral and non-neutral sequence evolution to search for genes in which the ratio of nonsynonymous to synonymous nucleotide substitutions (*d*N/*d*S, or *ω*) is higher in specified groups of eusocial lineages. Test 1 identified genes in which *ω* is higher in all eusocial lineages as a group relative to non-eusocial lineages, and did not discriminate between the highly and primitively eusocial lineages. Test 2 did so discriminate, and identified genes in which *ω* is highest in either all primitively or highly eusocial lineages as a group relative to all other lineages. These tests are not mutually exclusive; a gene may be evolving more rapidly in all eusocial relative to non-eusocial lineages, as well as evolving most rapidly in either the highly or primitively eusocial lineages.

The tests of heterogeneous rates of protein evolution yielded a number of genes evolving differently between eusocial and non-eusocial lineages, and among eusocial lineages. For Test 1, I found 212 out of 3638 genes (6%) evolving significantly more rapidly in all eusocial lineages relative to non-eusocial lineages ("All Eusocial Gene List"). For Test 2, I found 173 genes (5%) evolving most rapidly in highly eusocial lineages ("Highly Eusocial Gene List"), and 218 genes (6%) in primitively eusocial lineages ("Primitively Eusocial Gene List"), relative to other lineages [False Discovery Rate (FDR) adjusted $P < 0.05$ in all three cases]. Table 2 shows the most significant genes (based on *P*-value) on each list. These results demonstrate that the pattern of genetic changes associated with eusocial evolution includes some common changes and some changes that are unique to the different eusocial lifestyles.

Evaluation of biases in data. I explored the results of the tests to search for potential biases related to nucleotide composition or EST sequence coverage (see Supplement). I used Spearman's rank correlation to determine if the following characteristics of the sequence data were correlated with the *P*-values from the LRTs: 1) average GC content at the 3rd position; 2) average overall GC content; 3) *transition/transversion* ratio (kappa); and 4) *dN* tree length. The gappiness of an alignment could

introduce potential biases in the results (23, 24), so I also looked for correlations between the *P*-values from the LRTs and two metrics to assess coverage in the alignments: 1) gap percent (gapPCT), or the sum of the number of gaps in each sequence in an alignment divided by the sum of the total number of sites in all the sequences in an alignment; and 2) an alignment quality score (described in Supplement). Only a few of these characteristics of the data were significantly correlated ($P < 0.05$) with the *P*-values of the LRTs, but all correlations were very weak (range of Spearman's rho = -0.1 - 0.06, for all tests) (see Supplement).

Biological processes evolving more rapidly in eusocial relative to non-eusocial bees. I performed Gene Ontology (GO) enrichment analyses based on orthology to *Drosophila melanogaster* to identify biological processes that were overrepresented on the All Eusocial, Highly Eusocial, and Primitively Eusocial Gene Lists. GO enrichment analysis accounts for the over-representation of categories present in our set of 3638 genes, but the under-representation of some categories in this set is one explanation for why these categories may not have been enriched in our gene lists (data not shown). "Gland development" and "cell surface receptor-linked signal transduction" were among the terms overrepresented exclusively in the All Eusocial Gene List ($P < 0.05$, all GO results; results in Supplement).

Carbohydrate metabolism-related categories were enriched in both the All Eusocial and Highly Eusocial Gene Lists, suggesting that these genes are evolving both more rapidly in eusocial relative to non-eusocial lineages, and also most rapidly in highly eusocial lineages (Fig. 3A). Fifteen of the 26 genes encoding glycolytic enzymes in the dataset showed evidence of accelerated evolution in one or both of these lists (Fig. 3B), including enzymes that play a key regulatory role (phosphofructokinase) or are involved in glycolytic flux (hexokinase, pyruvate kinase). Subsequent analyses (see *Robustness of Results* below) revealed that seven out of these fifteen genes appear to be evolving most rapidly in honey bees

(genus *Apis*) (see Supplement). Two of the most rapidly evolving genes on the Highly Eusocial Gene List encode glycolytic proteins (Table 2).

Transcription-related categories were enriched in both the All Eusocial and Primitively Eusocial Gene Lists, but not in the Highly Eusocial Gene List. This suggests a similar pattern as was seen with carbohydrate-metabolism related genes in the All Eusocial and Highly Eusocial Gene Lists, only here with an emphasis in primitively eusocial lineages.

Lifestyle- and lineage-specific patterns of molecular evolution. Some biological processes were enriched exclusively in either the Highly Eusocial or Primitively Eusocial Gene Lists, and were not enriched in the All Eusocial Gene List (see Supplement). For example, I detected a signature of accelerated evolution in brain-related functional categories in primitively eusocial bees, but not in highly eusocial bees.

I performed an additional series of "lineage-specific" tests to identify genes evolving more rapidly in any individual eusocial lineages relative to all other lineages in the study (see Supplement). I was specifically interested in whether lineages with the same eusocial lifestyle showed similar biological processes undergoing accelerated evolution, but via changes in unique sets of genes. I did find evidence for this pattern in some lineages. For example, genes related to reproduction are rapidly evolving in both primitively eusocial lineages, Bombini and *E. robusta*, relative to all other lineages, but the actual genes in Bombini and *E. robusta* are largely different (data not shown).

Robustness of results. I performed an additional set of analyses to explore whether specific lineages may have contributed disproportionately to some of the results reported above. I performed "exclusion tests" in which I removed eusocial lineages from our alignments, one at a time, and re-ran Tests 1 and 2 to look for genes for which one species may have driven the pattern of accelerated evolution we

detected previously (see Supplement). Given that the removal of lineages can also affect statistical power to detect accelerated evolution in a gene (25), I consider this analysis to be useful for highlighting our strongest results, but do not believe this analysis is sufficient to invalidate the results obtained using the full set of species. I created three new gene lists by removing the genes from the original All, Highly and Primitively Eusocial Gene Lists whose significance appeared to have been driven by one eusocial lineage (see Supplement). GO enrichment analysis revealed that some of the trends identified in this analysis using all species were not robust to the removal of lineages (see Supplement), including the enrichment of "gland development" in the All Eusocial Gene List and the enrichment of transcription-related categories in the All Eusocial and Primitively Eusocial Gene Lists. Many biological processes were robust to the removal of lineages, including "cell surface receptor-linked signal transduction" in the All Eusocial Gene List, carbohydrate metabolism-related categories in the All Eusocial and Highly Eusocial Gene Lists, and neuron differentiation-related categories in the Primitively Eusocial Gene List.

I performed an additional analysis to determine whether artificial groupings of species would lead to the same enriched biological processes as the groupings of eusocial and non-eusocial lineages (see Supplement). GO terms were enriched in the lists of significant genes derived from each artificial grouping (data not shown), but these terms were largely different from those obtained from our eusocial groupings. This provides additional support that these results truly relate to eusocial evolution.

Discussion

I identified several hundred genes with a molecular signature of accelerated evolution, some with a signature in all eusocial lineages in our study, and some with a signature that was specific to a certain type of eusocial lifestyle or specific to individual lineages. Together, these results demonstrate that convergent evolution can involve a mosaic pattern of molecular changes in both shared and lineage-specific sets of genes.

Genes involved in gland development, signal transduction, and carbohydrate metabolism were among the most rapidly evolving genes identified in this study. These findings provide a starting point for linking specific genetic changes to the evolution of eusociality in bees, which will be an important challenge for the future. Major steps in this endeavor include: determining the consequences of the changes in amino acid sequence for protein function; learning how changes in protein function affect a particular biological process; and then understanding how evolutionary changes in a particular biological process might affect traits associated with eusociality (26). Below, I provide some speculation for the biological processes highlighted in our findings.

Genes associated with gland development appear to have been a strong target of selection during eusocial bee evolution. This is not surprising, because, relative to solitary insects, eusocial insects have remarkably diverse exocrine gland functions and produce many novel glandular secretions, including pheromones, brood food and antimicrobial compounds (8, 13, 14). Chemical signaling is a vital mechanism used to coordinate the behavior and physiology of colony members, and it is possible that at least some of the protein-coding sequence changes identified here are related to the evolution of advanced systems of chemical communication found in social bees.

Another category of genes that appear to have been a strong target of selection during eusocial bee evolution is genes involved in signal transduction. Signal transduction has been implicated in the regulation of behavior across disparate taxa (27), and several genes on the All Eusocial Gene List in this category have known roles in behavior and neuronal function, including *metabotropic GABA-B receptor subtype 1* (28). My results provide further evidence that signal transduction may be a general target of selection during behavioral evolution.

Genes associated with carbohydrate metabolism appear to have been a particularly strong target of selection during eusocial bee evolution. The finding of a shared pattern of accelerated evolution across all eusocial lineages in our study may reflect the fact that many eusocial bees rely more

heavily on highly processed honeys in their diet than do non-eusocial species (8), although all bees use nectar as their carbohydrate source. In addition, several characteristics shared by all eusocial insects, including worker-queen caste determination and worker-worker division of labor, are influenced by nutrition (6). Transcriptomic analyses have implicated highly conserved molecular pathways associated with metabolism (6), especially insulin signaling (29-31), in both the evolution and function of these traits (6, 32). My results are consistent with these findings, and further suggest that coding sequence changes in carbohydrate metabolism-related genes may have been involved in the evolution of these novel eusocial traits in bees.

Additional changes in carbohydrate metabolism-related genes were also detected in the highly eusocial bee lineages, and not in primitively eusocial bee lineages. This result may be due to the evolution of unique metabolic demands in the highly eusocial lifestyle, such as year-round nest thermoregulation (8), extended lifespan in queens (10-fold longer than workers) (13, 33), and greatly increased foraging activity (14). Nearly half of the genes in the glycolysis pathway that were enriched in the All Eusocial and/or Highly Eusocial Gene Lists were not robust solely to the removal of the *Apis* lineage from the analysis, suggesting that an abundance of changes in the glycolysis pathway may have occurred in this lineage.

I was initially surprised to detect a signature of accelerated evolution in brain-related GO categories in primitively eusocial bee lineages, but not in highly eusocial bee lineages. The Social Brain Hypothesis, developed to explain primate brain evolution, posits that the cognitive demands of social life are a strong selective force in brain evolution (34). It might be assumed that these demands are greater in the larger and more complex colonies of the highly eusocial bees, and thus a stronger signature of rapid evolution in brain-related genes would be found in highly eusocial relative to primitively eusocial lineages (35). However, perhaps it is the primitively eusocial society members that face greater sociocognitive challenges, because social roles are more fluid and the balance between

cooperation and competition is more dynamic in primitively eusocial colonies relative to the more structured highly eusocial colonies (8, 13, 35).

One rapidly evolving gene in the Primitively Eusocial Gene List, *dunce*, was originally identified as a *Drosophila* learning and memory mutant, and it has emerged as an important gene in the regulation of neural plasticity in both invertebrates and vertebrates (36). Recent studies implicate *dunce* and other genes in the cAMP pathway in social learning (37). Both the lineage-specific and robustness analyses suggest that, of the taxa studied here, *dunce* is evolving most rapidly in bumble bees. This finding of accelerated evolution in brain-related genes exclusively in primitively eusocial bees might eventually help us understand more about the evolution of behavioral differences that exist between primitively and highly eusocial species.

In addition to positive natural selection, non-adaptive phenomena such as relaxed constraint may contribute to the pattern of heterogeneous nucleotide substitution among sequences that was observed (16). Whether a gene is exposed to increased positive selection in eusocial lineages, or less purifying selection relative to non-eusocial lineages, is a distinction that I cannot formally establish. In both cases, a difference in selective regime between the eusocial and non-eusocial lineages resulted in an increased rate of protein evolution in the eusocial lineages. Other issues have been raised regarding the reliability of the statistical methods I used for detecting adaptive molecular changes in individual genes (38-40). However, my focus on identifying biological processes represented by groups of genes, rather than individual genes, ameliorates these concerns. It is unlikely that so many genes in a single functional GO category, particularly those involved in basic "housekeeping" processes (*e.g.*, carbohydrate metabolism), have been under relaxed constraint or exhibit consistent model departure stratified by sociality across lineages. The results I present motivate further investigation into differences in the functioning of these biological processes between eusocial and non-eusocial species, and the functional effects of the specific genetic changes identified.

A key finding in this study is that convergent evolution of eusociality in bees involves both shared and lineage-specific sets of genes. The lineage-specific findings suggest that the multiple, independent evolutionary paths to eusociality may have each been shaped by different combinations of extrinsic and intrinsic factors, and perhaps also via different forces of selection. In the future, it may be possible to use molecular signatures of selection on different functional classes of genes to identify which forces of selection were important in eusocial evolution. Recent evidence suggests that reproductive protein evolution can be driven by sexual selection (7), but it is not yet known if there are similar connections between other selective forces and functional classes of genes.

My finding of shared sets of rapidly evolving genes across three independent lineages that gave rise to eusociality in bees suggests that there might also be some common molecular roots for eusocial evolution, despite the incredible social diversity found among bees. Among the biological processes that appear to have been under selection across all eusocial lineages in this study, carbohydrate metabolism stands out. Insulin signaling, which is involved in carbohydrate metabolism, has been broadly implicated in the regulation of several eusocial traits, as mentioned above (6). It has been previously suggested that there is a "genetic toolkit" for eusociality, a set of highly conserved genes and molecular pathways that were co-opted for novel, social functions during eusocial evolution (32). My results provide additional support for the possibility that genes related to carbohydrate metabolism are key components of this putative toolkit (6, 32). The existence of a genetic toolkit for eusociality can be rigorously tested because there are at least another eight independent gains of eusociality in the bees, ants, wasps, and termites (2). The insect societies provide rich material to explore how changes in DNA sequence are associated with the evolution of social life.

Acknowledgments

I thank G. Bloch, C. Chanchao, K. Kapheim, M. de Lama, K. Hartfelder, M.P. Schwarz, T. Pitts-Singer, and S. Wongsiri for bee samples; E. Hadley for help with figures; T. Newman for laboratory assistance; C.

Rassmussen for field assistance and bee identification; T. O'Conner for sequence editing; B. Smith for bioinformatic assistance; J. Huelsenbeck, A. Meade and L. Jermiin for substitution model discussion; T.D. Seeley for helpful discussion; members of the Clark lab for helpful suggestions on the analyses; S.A. Ament, A.L. Toth, J.B. Whitfield, the Social Insect Reading Group (UIUC), and members of the Robinson and Cameron labs for comments on the manuscript; and P. Bouffard, T. Harkins, T. Jarvie, R. Martone and J. Osterberger (Roche/454) for expert technical support. Funding for this project came from 454 Life Sciences (Roche Diagnostics Corporation) via the Roche 1GB contest (G.E.R., S.A.C., A.G.C., M.E.H. & S.S.); NSF DEB07-43154 (G.E.R. and M.E.H.); and an NSF Predoctoral Fellowship (B.J.F.) and an NIH Training Grant (S.H.W.).

FIGURES AND TABLES

Figure 2. Bee species and evolutionary models used to identify genes evolving rapidly in eusocial lineages. (*A*) Phylogeny of species in study based on previously published trees (9–11) and EST data (see Supplement). Some analyses of EST data yielded an alternative topology; molecular evolutionary analyses performed with each topology gave highly similar results (see Supplement). Diamonds represent independent origins of eusociality. Reconstruction of eusocial origins based on phylogenies with greater taxon sampling (refs. 9–11; green dashed branches indicate position of non-eusocial lineages not included in the study). Lineages are color-coded by lifestyle: red, highly eusocial; blue, primitively eusocial; and green, non-eusocial. Boxes list key characteristics of each lifestyle (8, 13). (*B* and *C*) Representation of branch models of nonneutral evolution that were compared with null models by using likelihood ratio tests (LRTs). Lineages are color-coded as in *A*, except in test 1, where "All Eusocial" lineages are coded in purple. *B*, Test 1: $\omega_{Eusocial} \neq \omega_{Non\text{-}eusocial}$; *C*, Test 2: $\omega_{Highly\ eusocial} \neq \omega_{Primitively\ eusocial} \neq \omega_{Non\text{-}eusocial}$.

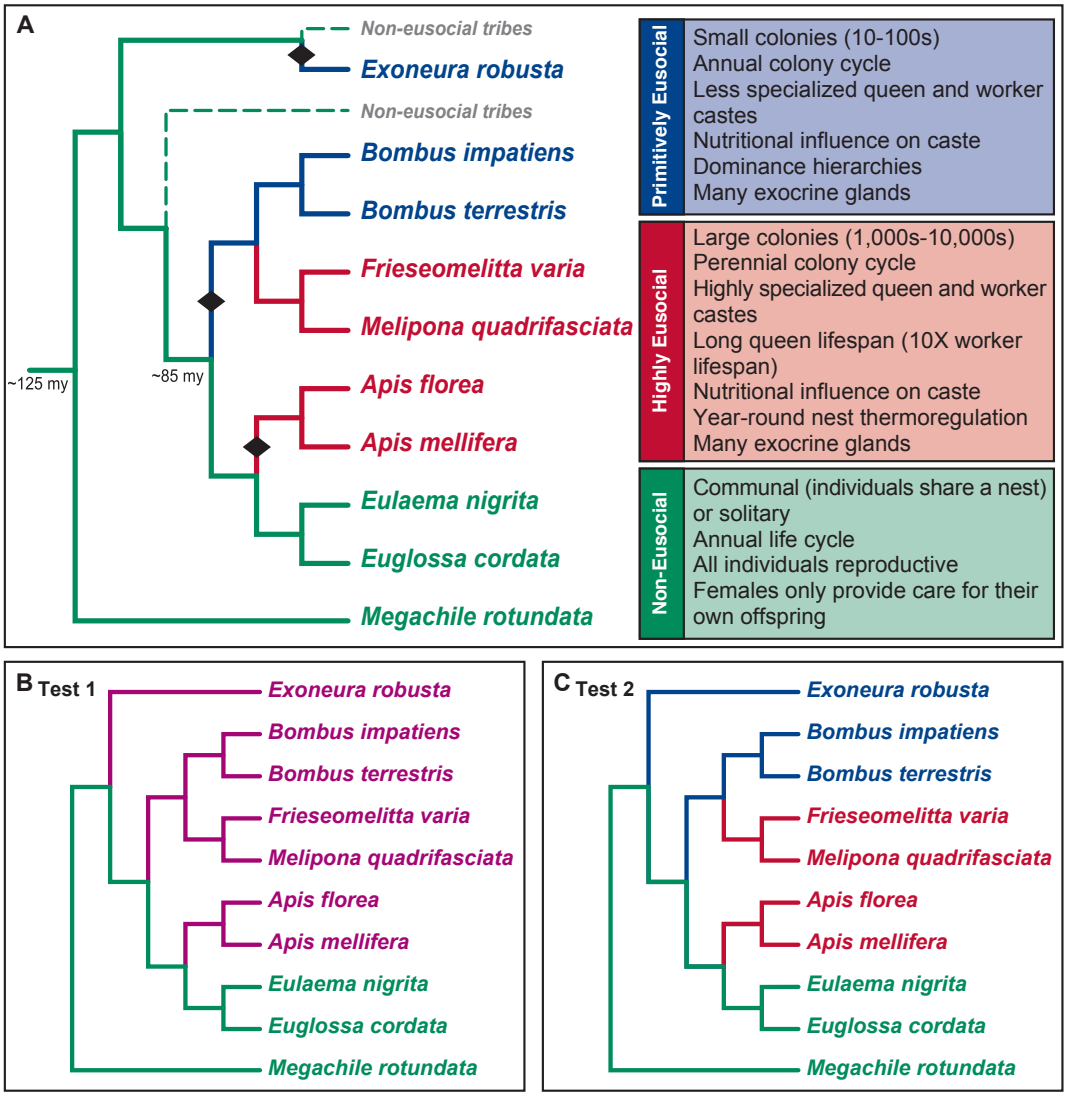


Figure 3. Biological processes with evidence of accelerated evolution in eusocial lineages. (*A*) Overlap of rapidly evolving biological processes based on GO enrichment analysis of the All Eusocial, Highly Eusocial, and Primitively Eusocial gene lists. Individual GO categories were condensed into "meta-categories" based on related function (see Supplement). (*B*) Rapidly evolving genes in the glycolysis pathway. *A. mellifera* genes mapped onto pathway based on orthology to *D. melanogaster* genes in KEGG (Kyoto Encyclopedia of Genes and Genomes, http://www.genome.jp/kegg/) glycolysis/gluconeogenesis pathway (dme00010). Genes likely evolving most rapidly in *Apis* are GB10695, GB10851, GB13401, GB16546, GB17113, GB17238, and GB19387.

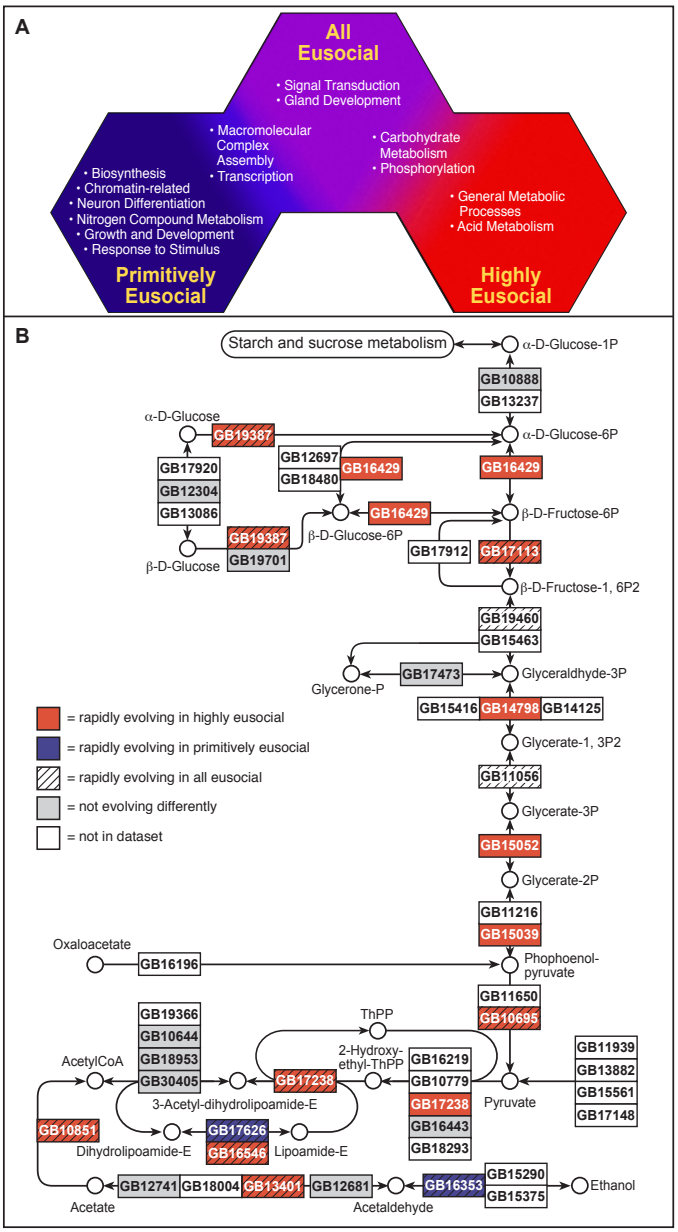


Table 2. Genes evolving more rapidly in eusocial bee lineages. Gene rank based on FDR-corrected *P*-values from likelihood ratio tests, in order of increasing significance. All genes in table were either robust against the removal of all lineages, or do not appear to be strongly driven by any one lineage (see Supplement). Relative ω is the fold difference compared to the non-eusocial ω.

Gene	Function	*A. mellifera* gene	Rank	*P*-value	Relative ω
Accelerated evolution in all eusocial lineages (Test 1)					
girdin	Actin-binding protein, involved in regulation of cell size	GB19383	6	0.00000	6.06
la autoantigen-like	Ribonucleoprotein, involved in development	GB14277	11	0.00015	3.51
brahma	Chromatin remodeler, involved in axonogenesis and oogenesis	GB30507	12	0.00015	4.25
syntaxin7	Membrane-bound protein involved in SNAP receptor activity	GB14433	15	0.00020	5.49
acetyl coenzyme A synthase	Enzyme, involved in metabolism of carbon sugars	GB10851	17	0.00027	3.40
Accelerated evolution in primitively eusocial lineages (Test 2)					
signal recognition particle 14 kDa	Involved in mRNA binding	GB15372	9	0.00000	136.5
no on or off transient A	mRNA binding protein, involved in courtship song in *Drosophila*	GB18173	10	0.00000	5.74
helicase 98B	Enzyme, involved in immune response	GB14810	11	0.00000	4.62
β spectrin	Cytoskeletal protein, involved in nervous system development	GB11407	12	0.00000	1.88
syntaxin7	Membrane-bound protein involved in SNAP receptor activity	GB14433	18	0.00000	7.64
Accelerated evolution in highly eusocial lineages (Test 2)					
phosphofructokinase	Enzyme involved in glycolysis	GB17113	3	0.00000	3.18
enolase	Enzyme involved in glycolysis	GB15039	4	0.00000	3.35
pelle	Serine/threonine kinase, involved in immune response and axon targeting	GB16397	5	0.00000	3.89
RhoGAP100F	GTPase, involved in axonogenesis and signal transduction	GB15150	25	0.00005	2.39
mitogen-activated protein kinase kinase 1	Protein kinase, involved in MAP kinase signaling	GB11819	34	0.00280	3.58

LITERATURE CITED

1. Wilson EO, Hölldobler B (2005) Eusociality: origin and consequences. *Proc Natl Acad Sci USA* 205:13367-13371.
2. Darwin C (1859) *On the Origin of Species* (Murray, London).
3. Queller DC (1992) A general model for kin selection. *Evolution* 46:376-380.
4. Nowak MA, Tarnita CE, Wilson EO (2010) The evolution of eusociality. *Nature* 466:1057–1062.
5. Smith CR, Toth AL, Suarez AV, Robinson GE (2008) Genetic and genomic analyses of the division of labour in insect societies. *Nat Rev Genet* 9:735-748.
6. Wong A (2010) Testing the effects of mating system variation on rates of molecular evolution in primates. *Evolution* 64:2779-2785.
7. Michener CD (1974) *The Social Behavior of the Bees* (Harvard Univ. Press, Cambridge, MA).
8. Cameron SA, Mardulyn P (2001) Multiple molecular data sets suggest independent origins of highly eusocial behavior in bees (Hymenoptera:Apinae). *Syst Biol* 50:194-214.
9. Schwarz MP, Richards MH, Danforth BN (2007) Changing paradigms in insect social evolution: insights from halictine and allodapine bees. *Annu Rev Entomol* 52:127-150.
10. Cardinal S, Straka J, Danforth BN (2010) Comprehensive phylogeny of apid bees reveals the evolutionary origins and antiquity of cleptoparasitism. *Proc Natl Acad Sci USA* 107:16207-16211.
11. Danforth BN, Sipes S, Fang J, Brady SG (2006) Recent and simultaneous origins of eusociality in halictid bees. *Proc Natl Acad Sci USA* 103:15118-15123.
12. Wilson EO (1971) *The Insect Societies* (Harvard Univ. Press, Cambridge, MA).
13. Roubik DW (1992) *Ecology and Natural History of Tropical Bees*. (Cambridge Univ. Press, Cambridge).
14. HBGSC (2006) Insights into social insects from the genome of the honeybee *Apis mellifera*. *Nature* 443:931-949.

15. Yang ZH, Bielawski JP (2000) Statistical methods for detecting molecular adaptation. *Trends Ecol Evol* 15:496–503.
16. Lein ES, et al. (2007) Genome-wide atlas of gene expression in the adult mouse brain. *Nature* 445:168-176.
17. Katoh K, Misawa K, Kuma K, Miyata T (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucl Acids Res* 30:3059-3066.
18. Drummond AJ, et al. (2010) Geneious v5.1, Available from http://www.geneious.com.
19. Ronquist F, Huelsenbeck JP (2001) MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* 19:1572-1574.
20. Yang ZH (2007) PAML 4: Phylogenetic analysis by maximum likelihood. *Mol Biol Evol* 24:1586-1591.
21. Huang DW, Sherman BT, Lempicki RA (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protoc* 4:44-57.
22. Schneider A, *et al*. (2010) Estimates of positive Darwinian selection are inflated by errors in sequencing, annotation, and alignment. *Genome Biology and Evol* 1:114-118.
23. Wong KM, Suchard MA, Huelsenbeck JP (2008) Alignment uncertainty and genomic analysis. *Science* 319:473-476.
24. Wong, WSW, Yang Z, Goldman N, Nielsen R (2004) Accuracy and power of statistical methods for detecting adaptive evolution in protein coding sequences and for identifying positively selected sites. *Genetics* 168: 1041-1051.
25. Dean AM, Thornton JW (2007) Mechanistic approaches to the study of evolution: the functional synthesis. *Nat Rev Genet* 8:675-688.
26. Ben-Shahar Y, Robichon A, Sokolowski MB, Robinson GE (2002) Influence of gene action across different time scales on behavior. *Science* 296:741-744.

27. Root CM, et al. (2008) A presynaptic gain control mechanism fine-tunes olfactory behavior. *Neuron* 59:311-321.
28. Ament SA, Corona M, Pollock HS, Robinson GE (2008) Insulin signaling is involved in the regulation of worker division of labor in honey bee colonies. *Proc Natl Acad Sci USA* 105:4226-4231.
29. Corona M, et al. (2007) Vitellogenin, juvenile hormone, insulin signaling, and queen honey bee longevity. *Proc Natl Acad Sci USA* 104:7128-7133.
30. Patel A, et al. (2007) The making of a queen: TOR pathway is a key player in diphenic caste development. *PLoS One* 2:e509.
31. Toth AL, Robinson GE (2009) Evo-devo and the evolution of social behavior: brain gene expression analyses in social insects. *Cold Spring Harb Symp Quant Biol* 74:419-426.
32. Finkel T, Holbrook N (2000) Oxidants, oxidative stress and the biology of ageing. *Nature* 405:239-247.
33. Dunbar RIM (2009) The social brain hypothesis and its implications for social evolution. *Ann Hum Biol* 36:562-572.
34. Gronenberg W, Riveros AJ (2009) In *Organization of Insect Societies: From Genome to Sociocomplexity*, eds Gadau J, Fewell J (Harvard University Press), pp 377-401.
35. Silva A, et al. (1998) CREB and memory. *Annu Rev Neurosci* 21:127-148.
36. Sokolowski MB (2010) Social interactions in ‘‘simple’’ model systems. *Neuron* 65:780-794.
37. Nozawa M, Suzuki Y, Nei M (2009) Reliabilities of identifying positive selection by the branch-site and the site-prediction methods. *Proc Natl Acad Sci USA* 106:6700–6705.
38. Suzuki Y, Nei M (2002) Simulation study of the reliability and robustness of the statistical methods for detecting positive selection at single amino acid sites. *Mol Biol Evol* 19:1865–1869.

39. Wolf JBW, Kunstner A, Nam K, Jakobsson M, Ellegren H (2009) Nonlinear dynamics of nonsynonymous (d(N)) and synonymous (d(S)) substitution rates affects inference of selection. *Genome Biol Evol* 1:308–319.

CHAPTER 4: SOCIAL REGULATION OF MATERNAL TRAITS IN NEST-FOUNDING BUMBLE BEE (*BOMBUS TERRESTRIS*) QUEENS

ABSTRACT

During the nest-founding phase of the bumble bee colony cycle, queens undergo striking changes in maternal care behavior. Early in the founding phase, prior to the emergence of the first workers in the nest, queens are reproductive and also provision and feed their offspring. However, later in the founding phase, queens cease feeding offspring and become specialized on reproduction. This transition is synchronized with the emergence of the first workers in the colony, who assume the task of feeding their siblings. Using a social manipulation experiment, I tested the hypothesis that workers socially regulate the transition from feeding brood to specialization on reproduction in nest-founding bumble bee queens. Consistent with this hypothesis, I found that early-stage queens with workers prematurely added to their nests reduce their brood-feeding behavior and increase egg-laying, and likewise, late-stage queens increase their brood-feeding behavior and decrease egg-laying when workers are removed from their nests. Further, brood-feeding and egg-laying behavior were negatively correlated in these queens. I used an Agilent brain EST-based microarray to explore a second hypothesis, that workers alter brain gene expression in nest-founding queens. I found evidence that brain gene expression in nest-founding queens is altered by the presence of workers, with the effect much stronger in late-stage founding queens. Additionally, expression levels of some genes were correlated with quantitative differences in brood-feeding and egg-laying behavior. This study provides new insights into how the transition from feeding brood to specialization on reproduction in bumble bee queens is regulated during the nest initiation phase of the colony cycle.

INTRODUCTION

Maternal care is a hallmark of many animal societies (1). In queens of the "primitively" eusocial bumble bees (genus *Bombus*, family Apidae), striking changes in maternal care behavior occur during the nest founding phase of the life cycle, when mated queens emerge from overwintering and found new nests (2). Early in the founding phase, prior to the emergence of workers in the nest, queens are reproductive and also provision their nests and progressively feed their larval offspring. However, later in the founding phase, following the emergence of workers in the nest, queens cease feeding offspring and instead become largely specialized on reproduction, as workers take over the task of feeding brood in the colony. These changes make bumble bee queens an excellent system for studying the regulation and plasticity of maternal care.

A common feature of animal societies is social regulation of the behavior and physiology of group members (1). Social regulation serves to promote cohesion between society members, for example by organizing reproductive efforts (3), reducing conflict (4), and coordinating the labor force (5). Various forms of social regulation have been demonstrated in bumble bee societies, including regulation of circadian rhythms in queens by brood (6), pheromonal regulation of worker foraging activity (7), and inhibition of worker reproduction by both queens (8) and workers (9).

I performed a social manipulation experiment to test the hypothesis that there is social (worker) regulation of the transition from feeding brood to specialization on reproduction in nest-founding bumble bee queens. This hypothesis was based on two previous findings in bumble bees: 1) the transition coincides with the emergence of workers in the colony (2); and 2) workers stimulate egg-laying behavior in nest-founding queens (10-12). Based on this hypothesis, I predicted that the artificial addition and removal of workers from recently founded nests (Fig. 4) would alter both maternal care and reproductive behaviors in nest-founding queens. I predicted that the artificial, premature addition of workers to nests in which workers had not yet emerged ("early-stage" nests) would cause a decrease

in brood-feeding behavior and an increase in egg-laying behavior in queens. I also predicted that removal of all workers from nests in which workers were beginning to emerge ("late-stage" nests) would cause an increase in brood-feeding behavior and a decrease in egg-laying behavior in queens.

I also explored the hypothesis that changes in brain gene expression are associated with the transition away from brood-feeding in nest-founding queens. In many cases of social regulation, social signals, which are initially received and processed by sensory and neural systems, ultimately lead to behavioral changes via changes in brain gene expression (13). To test this hypothesis, I studied patterns of brain gene expression in early- and late-stage founding queens with and without workers (Fig. 4) using an Agilent microarray made specifically for the bumble bee used in this study, *Bombus terrestris*, which was based on a large-scale brain transcriptomic project (14). I predicted that some genes would be differentially expressed in the brains of founding queens with *versus* without workers, which may represent components of the transcriptional architecture underlying social regulation of maternal care and reproduction in bumble bee queens. Understanding the ways in which social signals are transduced via transcriptional mechanisms is important because these mechanisms can provide important insights into how sociality evolved at the molecular level (15).

METHODS

Bees. B. terrestris is a well-studied (16, 17), temperate bumble bee that is commercially managed for pollination services (18). 47 recently initiated (*i.e.,* eggs present but no larvae) nests were brought into the laboratory and kept under conditions described in (16, 17). To synchronize the treatment groups, 22 of the 47 nests were brought to the laboratory first and allowed to develop until 2 d after the first workers emerged (~20 d); queens in these nests were "late-stage" founding queens (LQ). Fourteen days later, 25 additional nests were brought into the laboratory and allowed to develop until larvae were

present (~5 d) but no workers had yet emerged; queens in these nests were "early-stage" founding queens (EQ).

Social manipulation. For 12 of the 25 EQ nests, callow workers (< 24 h old) were added to simulate premature emergence of workers ("EQ+W" group). For the remaining 13 EQ nests, workers were not added and thus these nests were without workers for the duration of the experiment ("EQ-W" group). For 11 of the 22 LQ nests, any workers that emerged were removed < 24 h after their emergence ("LQ-W" group). For the other 11 LQ nests, any workers that emerged were removed < 24 h post-emergence and replaced with callow workers as a sham control ("LQ+W" group). To maintain the same number of workers in all "+W" nests, the following numbers of workers were added to EQ+W and LQ+W nests on days 1-7: 1, 2, 1, 1, 0, 0, 0; these numbers were based on the average number of workers to emerged in a subset of colonies (data not shown). All callow workers added to nests originated from the same source colonies to control for potential effects of worker genotype.

Observations of maternal care behavior. 42 of the 47 nest-founding queens were observed for 10 min periods on days 3, 5, and 7 after the first day workers were removed or added to the nests (day 0), for a total of 30 min observation per nest across the three days. The five queens not observed were from the EQ+W (n = 2) and EQ-W (n = 3) groups; data on reproductive behavior, colony development, and brain gene expression (described below) from these queens were included in other analyses. During each observation period the queens were continuously observed and the occurrence of any brood feeding event was recorded. Brood feeding is an easily observable, discrete behavioral sequence that lasts approximately 5-15 seconds. Larvae are clumped together spatially, and when bees feed brood, they open and regurgitate food into the larval cells, which can be observed when a bee places her mouthparts into an open cell and her abdomen contracts (2).

Reproductive behavior and colony development. At the end of day 7, all nests were collected and stored at -4˚C until nest dissections occurred. The number of eggs in the nest at the end of the experiment was recorded as a proxy for queen reproductive behavior; it is unlikely that any eggs were laid by workers, as the nests were very young and queens are highly effective at policing worker reproduction, even when nests are much older and larger (19, 20). The numbers of larvae and pupae present in the nests at the end of the experiment were also recorded as measures of colony development.

Statistical tests of maternal care and reproductive behavior. Two-tailed *t*-tests were used to compare the mean numbers of brood feeding events performed by queens in the four groups on each day of observation (days 3, 5, and 7) as well as the number of eggs, larvae, and pupae in the nests. To test for association between maternal care and reproductive behaviors in the queens, pairwise Pearson correlation coefficients were calculated between the total number of feeding events performed across the 3 days of observation and the number of eggs and larvae in the nest, using data from the 42 queens that were observed.

RNA preparation. At the end of day 7, all bees were collected within 1 h to control for circadian effects on gene expression. The bees were collected directly onto dry ice and their heads were immediately removed, placed in liquid Nitrogen, and stored at -80˚C to ensure RNA preservation. For brain dissections, whole bee heads were partially lyophilized and dissections were performed over dry ice (21). RNA was isolated from dissected brains using an RNeasy Mini Kit (QIAGEN) following the kit protocol except that the initial homogenization was done in a 500 ul microfuge tube using 100 ul extraction buffer.

Microarrays. An Agilent 4 x 44K *Bombus terrestris* Brain EST microarray was designed using sequence

data obtained from 454/Roche pyrosequencing of *B. terrestris* mRNA derived primarily from brain tissue (90% and 10% abdomen) (14). 250-1000 ng total RNA per sample was reverse transcribed and linear amplified according to manufacturer's instructions (Agilent Technologies). Samples (one per array) were hybridized on the microarray slide and washed according to the Agilent protocol. Slides were scanned using an Axon 4000B scanner and images were analyzed with GENEPIX software (Agilent Technologies, Santa Clara, CA). Methods described in greater detail in (22, 23).

Microarray data pre-processing and statistical analyses were done in R (24) using the limma package (25). Median foreground and median background values from the 47 .gpr files were read into R and any spots manually flagged (-100 values) were given a weight of zero. Background values were ignored because the use of these values to adjust for background fluorescence added more noise to the data, and the background was low and even for all arrays. Expression values were normalized with the quantile method (26) then log2 transformed. Coefficient of variation (CV) values were calculated across all 47 samples; the distribution of CV values was strongly bimodal, separating at a value of 0.015 (data not shown), suggesting a subset of oligos with little variation overall and hence no useful information; nearly all of the positive and negative control spots had low CV values as well. All control spots and oligos with CV values < 0.015 were removed from subsequent analyses, leaving 36,869 spots out of 45,220.

Because the microarray was designed from EST sequence rather than genomic data, stringent annotation criteria based on orthology to the honey bee (*Apis mellifera*) genome (27) were used to minimize the chances of including oligos that did not represent true *B. terrestris* genes in the subsequent statistical analyses. *A. mellifera* orthologs were identified using BLAST to the non-redundant nucleotide database in GenBank; annotated oligos had strongly supported (E-value < 0.0001) hits to genes in the Official Honey Bee Gene Set Version 2 (27). A total of 9924 oligos (putatively representing 4468 unique *A. mellifera* genes) met the annotation criteria and were included in subsequent analyses.

ANOVA. To identify genes associated with presence of workers and with stage of the queen (early *vs.* late), a 2*x*2 factorial model (24) fit in limma (25) was used to compare expression values across the four groups, taking into account which slide the array was on. Using ANOVA, the main effects of presence of workers and stage of the queen and the interaction term were estimated, as well as mean expression levels for the four groups. Multiple-test adjustment was performed using the False Discovery Rate (FDR) method (28) and probes with FDR $P < 0.05$ were considered significant.

Pairwise tests. Four pairwise tests (referred to hereafter as "Pairwise Tests") were performed to identify genes differentially expressed in the following contrasts: (*i*) EQ+W *vs.* EQ-W, (*ii*) LQ+W *vs.* LQ-W, (*iii*) LQ+W *vs.* EQ+W, and (*iv*) LQ-W *vs.* EQ-W. Multiple test adjustment was done separately for each contrast using the FDR method (28) and probes with FDR $P < 0.05$ were considered significant.

Rank Product Tests. An additional, non-parametric analysis based on the Rank Product test (referred to hereafter as "Rank Product Tests") was performed using the RankProd package (29, 30). The Rank Product test ranks oligos in each replicate according to expression value, with the underlying assumption that oligos that change expression level will change in rank. This additional analysis was used because of the small number of significant oligos in the ANOVA and Pairwise Tests, and also because within-group variation in expression values was high for many oligos. A rank sum analysis, which compares the sums of the ranks between two groups, was used because it requires greater consistency between replicates; the large numbers of replicates within the four groups (n = 11-13) led to spurious results using the more traditional rank product analysis. To correct for the effect of which slide the array was on, all normalized expression values from all treatment groups were batch-corrected using the removeBatchEffect function.

RankProd tests up- and down-regulation separately and corrects for multiple hypothesis testing

using a permutation-based approach (n = 100) to calculate the percentage of false-positives (PFP), which is an estimate of the false discovery rate (30). Therefore, the four pairwise comparisons between treatment groups yielded eight significant (PFP < 0.05) Rank Product Test gene lists: (*i*) upregulated in EQ+W *vs.* EQ-W; (*ii*) downregulated in EQ+W *vs.* EQ-W; (*iii*) upregulated in LQ+W *vs.* LQ-W; (*iv*) downregulated in LQ+W *vs.* LQ-W; (*v*) upregulated in LQ+W *vs.* EQ+W; (*vi*) downregulated in LQ+W *vs.* EQ+W; (*vii*) upregulated in LQ-W *vs.* EQ-W; and (*viii*) downregulated in LQ-W *vs.* EQ-W.

Associations between brain gene expression and behavior. Using data from the 42 observed queens, Pearson correlation coefficients were calculated between the normalized expression values for all 9924 annotated oligos on the microarray and the total number of brood feeding events observed on day 7, as well as between expression values and the number of eggs in the nest. Because no genes were significantly correlated with behavioral data at FDR $P < 0.05$, results reported for this test are significant at an uncorrected $P < 0.001$.

Functional analysis of gene lists. Gene Ontology (GO) enrichment analysis was performed on all gene lists of sufficient size using the Gene Ontology (GO) functional annotation tool (GOFat level) on the DAVID website (31). For better comparison to the Rank Product Test gene lists, each of the 4 Pairwise Test gene lists was separated into upregulated and downregulated lists for GO enrichment analysis. All gene lists were transformed into lists of *Drosophila melanogaster* orthologs using a previously published *A. mellifera* - *D. melanogaster* ortholog list (27). Results reported are *biological process* and *molecular function* terms enriched within the lists of *D. melanogaster* orthologs at an uncorrected $P < 0.05$, with terms with < 5 genes excluded. For the background list used for all tests of enrichment, *D. melanogaster* orthologs (n = 3369) were identified for all 9924 annotated probes on the *B. terrestris* microarray.

RESULTS

Effect of treatment on maternal care behavior. Artificial manipulation of the social environment of nest-founding queens had a strong effect on maternal care behavior (Fig. 5A; Tables A1-A3, Appendix A). On the first day of observation (day 3), EQ-W performed more brood feeding than EQ+W ($P < 0.05$), and this difference became greater on days 5 and 7 ($P < 0.001$, both days). There was no difference in brood feeding behavior in LQ-W and LQ+W on day 3 ($P = 0.06$), but by days 5 and 7, LQ-W did perform more brood feeding than LQ+W ($P < 0.01$, both days).

Differences in brood feeding behavior between queens of the same stage with *vs.* without workers could not be attributed to differences in amounts of larvae in the nests, as these numbers were similar in EQ+W and EQ-W nests ($P = 0.2$) and in LQ+W and LQ-W nests ($P = 0.8$) (Fig. 6A). Whereas there was no difference in the number of pupae in EQ+W and EQ-W nests ($P = 0.7$), there were more pupae in LQ+W nests relative to LQ-W nests ($P < 0.05$).

EQ and LQ with the same social conditions performed the same amount of brood feeding on days 3, 5, or 7 (+W nests, $P = 0.1$, 1.0, and 0.5, respectively; -W nests, $P = 0.8$, 0.1, and 0.6, respectively). The similarity in brood feeding behavior in EQ and LQ with the same social conditions existed despite the fact that LQ nests had nearly four times the number of larvae as EQ nests ($P < 0.001$) (Fig. 6A).

Effect of treatment on reproductive behavior. Artificial manipulation of the social environment also had a strong effect on reproductive behavior in nest-founding queens (Fig. 5B; Tables A1-A3, Appendix A). EQ+W nests had more than three times the number of eggs as EQ-W nests ($P < 0.01$). Likewise, LQ+W nests had more than four times the number of eggs as LQ-W nests ($P < 0.001$). LQ-W and EQ-W nests had similar amounts of eggs ($P = 0.3$), whereas LQ+W nests had more eggs than EQ+W nests ($P < 0.05$).

Association between maternal care and reproductive behavior. Across the 42 observed queens, there was a significant, negative, correlation (P = 0.002; r = -0.47) between the total number of brood feeding events performed and number of eggs in the nest (Fig. 5C). There was no significant correlation between the total number of brood feeding events performed and number of larvae in the nest (P = 0.5), or between number of larvae and number of eggs in the nest (P = 0.3).

Gene Expression: ANOVA. Table 3 shows the numbers of oligos and corresponding genes that were significantly differentially expressed in the ANOVA and also in the Pairwise and Rank Product Tests. For the ANOVA, four oligos (corresponding to four genes) were differentially expressed in queen brains in association with the presence of workers. Included in this list were *aldehyde dehydrogenase type III*, which encodes an enzyme involved in the oxidation of aldehydes, and *arrestin 2*, a gene involved in phototransduction in bees (32) as well as olfactory and other types of sensory transduction in other animals (33).

394 oligos (corresponding to 328 genes) were differentially expressed in queen brains in association with the stage of the queen. This list contained several genes previously identified as important in insect behavior, including *calcium/calmodulin-dependent protein kinase*, a gene involved in learning and memory in flies (34); *inositol 1,4,5,-tris-phosphate receptor*, a gene that affects feeding- and flight-related behaviors in *Drosophila* (35, 36) and also has patterns of expression associated with behavioral maturation in honey bees (37, 38) and nest-provisioning behavior in *Polistes* wasps (39); and *nejire*, a CREB-binding protein and a transcriptional coactivator involved in the regulation of circadian rhythms in *Drosophila* (40). *nejire* was shown to be under positive natural selection in multiple bumble bee species (including *B. terrestris*) relative to several other eusocial and non-eusocial bee lineages (14). The biological processes *cognition*, *behavior,* and *neurological system process* were among the terms enriched in this gene list (Table 4).

Gene Expression: Pairwise Tests. No oligos were differentially expressed between the EQ+W and EQ-W groups. 37 oligos (corresponding to 37 genes) were differentially expressed between the LQ+W and LQ-W groups. Included in this list was a gene involved in circadian activity, *cycle*, whose expression has previously been shown to be socially regulated in honey bees (41), and an NDMA receptor (*NDMA receptor 1*). These results are consistent with those from the ANOVA, which revealed only four genes to be differentially expressed in association with the presence of workers.

1101 oligos (corresponding to 884 genes) were differentially expressed between the LQ-W and EQ-W groups. This list contained additional genes involved in inositol signaling, including *inositol-3-phosphate synthase, no receptor potential A*, and *synaptojanin*, and several genes involved in aging, including *target of rapamycin*, a gene in the insulin signaling pathway that is involved in the regulation of behavioral maturation and aging (42-44) in honey bees. The functional terms enriched in the lists of genes up- and down-regulated in LQ-W *vs.* EQ-W are provided in Table 5. No oligos were differentially expressed between the LQ+W and EQ+W groups.

Rank Product Tests. Nine oligos (corresponding to eight genes) were upregulated and no oligos were downregulated in the EQ+W group relative to the EQ-W group using the Rank Product test. 338 oligos (corresponding to 304 genes) were upregulated in the LQ+W group relative to the LQ-W group, including oligos corresponding to the genes *cycle* and *NDMA receptor 1*. 149 oligos (corresponding to 132 genes) were downregulated in the LQ+W group relative to the LQ-W group.

39 oligos (corresponding to 28 genes) were upregulated in the LQ+W group relative to the EQ+W group and 21 oligos (corresponding to 18 genes) were downregulated in the LQ+W group relative to the EQ+W group. 575 oligos (corresponding to 453 genes) were upregulated in the LQ-W group relative to the EQ-W group, including oligos corresponding to the genes *inositol-3-phosphate synthase, no receptor potential A,* and *target of rapamycin.* The functional terms enriched in this gene list included

several terms related to cellular development and death, including *apoptosis* (Table 6). 685 oligos (corresponding to 566 genes) were downregulated in the LQ-W group relative to the EQ-W group, including an oligo corresponding to *foraging*, a gene involved in the regulation of feeding- and foraging-related behaviors in honey bees and other species (45). Multiple terms related to memory were enriched within this gene list (Table 6). Figure 7 shows a comparison of the genes significant for the Pairwise and Rank Product Tests.

Associations between brain gene expression and behavior. Across the 42 observed queens, 17 oligos (corresponding to 17 genes) had expression levels that were significantly correlated with the total number of brood-feeding events performed on day 7 and eight oligos (corresponding to eight genes) had levels of expression significantly correlated with number of eggs in the nest (Table 7).

DISCUSSION

I used a social manipulation experiment to provide evidence that workers socially regulate queen maternal traits during the nest initiation phase of the bumble bee *B. terrestris*. The transition from feeding brood to specialization on reproduction in nest-founding queens appears to be highly plastic in both directions. Depending on social conditions, queens can increase brood feeding and decrease egg laying, beyond the point when workers emerge in the nest, and likewise, queens can decrease brood feeding and increase egg laying, prior to the natural emergence of workers in the nest. Two other major transitions in the bumble bee queen life cycle also appear to be socially regulated by workers, the "switch point," where queens switch from laying diploid (*i.e.,* female-destined) to haploid (*i.e.,* male-destined) eggs, and later on the "competition point," which is typically followed by a dramatic increase in queen-worker conflict in the nest (46).

Despite exhibiting considerable behavioral plasticity, there appear to be limits to the amounts of brood feeding and reproduction that queens can perform. Although the nests of late-stage founding queens without workers had far more larvae than nests of early-stage queens without workers, queens in these nests performed similar amounts of brood feeding, suggesting that a maximum rate of brood feeding may have been reached. Additionally, the negative association between brood feeding and reproduction across queens in all four groups suggests that a trade-off may exist, perhaps due to constraints stemming from physiological, pleiotropic, or other connections between the two behaviors.

Surprisingly, despite the strong effect of the treatment on behavior, the number of genes associated with the presence of workers in the ANOVA was relatively small compared to other studies on the social regulation of gene expression, many of which used similar methods to measure expression (22, 47-50). In addition, the effect of workers on queen brain gene expression was highly dependent on the stage of the queen, as was revealed by the Pairwise and Rank Product Tests. In both sets of tests, far more genes were differentially expressed between late-stage queens that varied in social environment relative to the early-stage queens. Given that worker loss occurs in the wild and is likely detrimental for nests, perhaps bumble bee queens have evolved the ability to respond to worker loss in a stereotyped way that involves changes in gene expression. Alternatively, the unnatural, premature addition of workers to early-stage nests may have precipitated a more "opportunistic" response by queens in these nests. An opportunistic response may not have required changes in brain gene expression, or may have involved expression changes that were not detectable at the time point that we sampled (13). In cichlid fish, dramatic behavioral changes that occur in response to changes in social environment precede detectable changes in gene expression (51), suggesting that gene expression changes are not necessary to initiate behavioral responses in this system.

The gene *cycle* is a strong candidate for future studies on the molecular basis of social regulation of queen maternal traits. *cycle*, which is a molecular component of the circadian clock, was differentially

expressed between late-stage queens with *versus* without workers in both the Pairwise and Rank Product Tests. This finding suggests that the social environment may alter queen behavior via changes in circadian gene expression. The finding that rhythmicity in nest-founding queens is influenced by social environment in a previous study on *B. terrestris* (5) is also consistent with this hypothesis.

In contrast to the small number of genes associated with the presence of workers, many genes were associated with stage of the queen in ANOVA. In terms of absolute time, the difference between early- and late-stage founding queens in the experiment was small (on the order of weeks) relative to their lifespan (2). Regardless, queens may undergo important, internal developmental changes during this critical period in the life cycle, which occur irrespective of the social environment or brood feeding and egg-laying behavior. The finding that genes involved in circadian rhythms, learning and memory, experience, and aging were differentially expressed between early- and late-stage queens is consistent with the idea that intrinsic developmental changes may be occurring in association with gene transcriptional changes during this transitional period in the queen life cycle.

Although I have attributed the differences in queen behavior and gene expression in this experiment to the social effect of workers, social signals originating from brood may have also influenced these traits, as the number of eggs and pupae differed between nests with and without workers. Previous studies in bumble bees suggest that pupae do influence egg-laying behavior in queens (10-12). However, the number of larvae in the nests of queens of the same stage did not differ in my study, suggesting that this factor did not influence queen behavior or gene expression.

A small number of genes on the microarray had levels of expression that were correlated with brood feeding or egg laying behavior across all queens in the experiment. Previous studies in fish (52) and monkeys (53) have demonstrated correlations between social behaviors and gene expression in brain tissue. The set of genes with levels of expression correlated with behavior in this experiment includes the gene *feminizer*, a gene involved in sex determination, as well as genes involved in glycolipid

metabolism and putative zinc finger proteins. These genes are an excellent set of candidates for future studies on the molecular regulation of maternal and reproductive behaviors in bumble bees.

With many wild bumble bee populations in decline (54), understanding the transition from feeding brood to specialization on reproduction in nest-founding queens may be important for predicting how bumble bees can respond to environmental changes. This includes understanding how the transition is regulated and what the limits of queen behavioral plasticity in queens of different ages. Although queens in this study responded to changes in social environment, the absence of workers was associated with a significant decline in queen reproductive output, suggesting that maintaining a substantial worker population is important for the viability of young colonies. Additionally, the results of this study suggest that artificially adding workers to young nests may be a valuable strategy for minimizing the loss of colonies reared for pollination, a finding that is consistent with previous studies (10-12).

ACKNOWLEDGMENTS

I thank Y. Gruber and M. Tamarkin for assistance with behavioral observations; H. Sphigler for input on experimental design; M. Cohen for technical assistance; G. Bloch for input on experimental design, interpretation of data, and input on the manuscript; and other members of the GB laboratory for input on experimental design. Thanks to M. Band and J. Drnevich at the Keck Center for Bioinformatics, University of Illinois at Urbana-Champaign, Yad Mordechai for bees, and K. Varala, M.E. Hudson, and A. Venkat for bioinformatic assistance. The study was supported by 454 Life Sciences (Roche Diagnostics Corporation) via the Roche 1GB contest, National Science Foundation Grant DEB07-43154 (to G.E. Robinson, Primary Investigator) and National Institutes of Health–University of Illinois Sensory Neuroscience Training Grant PHS2T32DC006612 (to S.H. Woodard; A.S. Feng, Principal Investigator).

DATA AQUISITION

The microarray results from this study can be accessed at the ArrayExpress website (http://www.ebi.ac.uk/arrayexpress/).

FIGURES AND TABLES

Figure 4. Experimental design to test for effects of workers and stage of queen. For Worker axis, "+" = workers present in nest, "-" = workers not present in nest. For Stage axis, "Early" = early-stage queen, "Late" = late-stage queen.

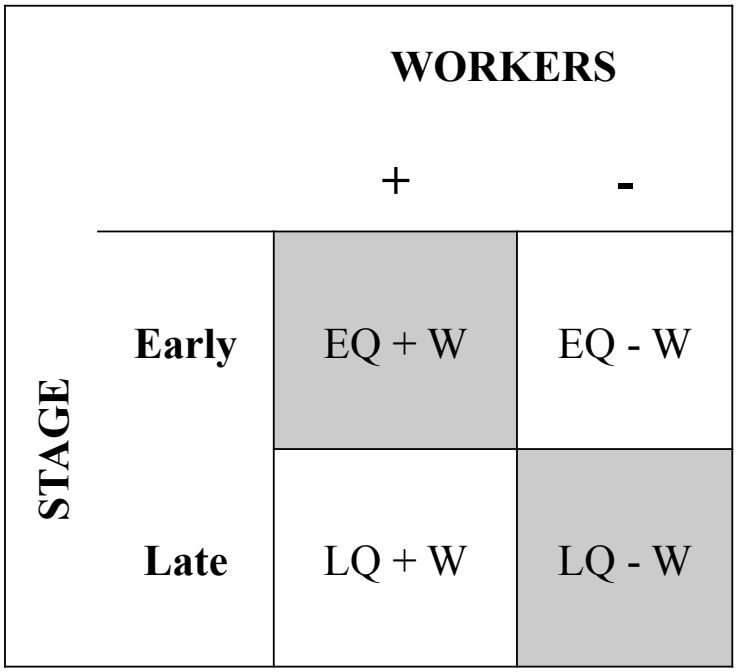


Figure 5. Effect of social manipulation on queen behavior. A) Effect of treatment on brood feeding behavior; *x*-axis = day of experiment, *y*-axis = number of brood feeding observed; values are group means with standard error bars. B) Effect of treatment on reproductive behavior; *x*-axis = group, *y*-axis = number of eggs in nests on day 7; values shown are group means with standard error bars; letters represent results of post-hoc pairwise *t*-tests. C) Association between brood feeding and reproductive behavior; *x*-axis = number of brood feeding observed, *y*-axis = number of eggs in nests on day 7.

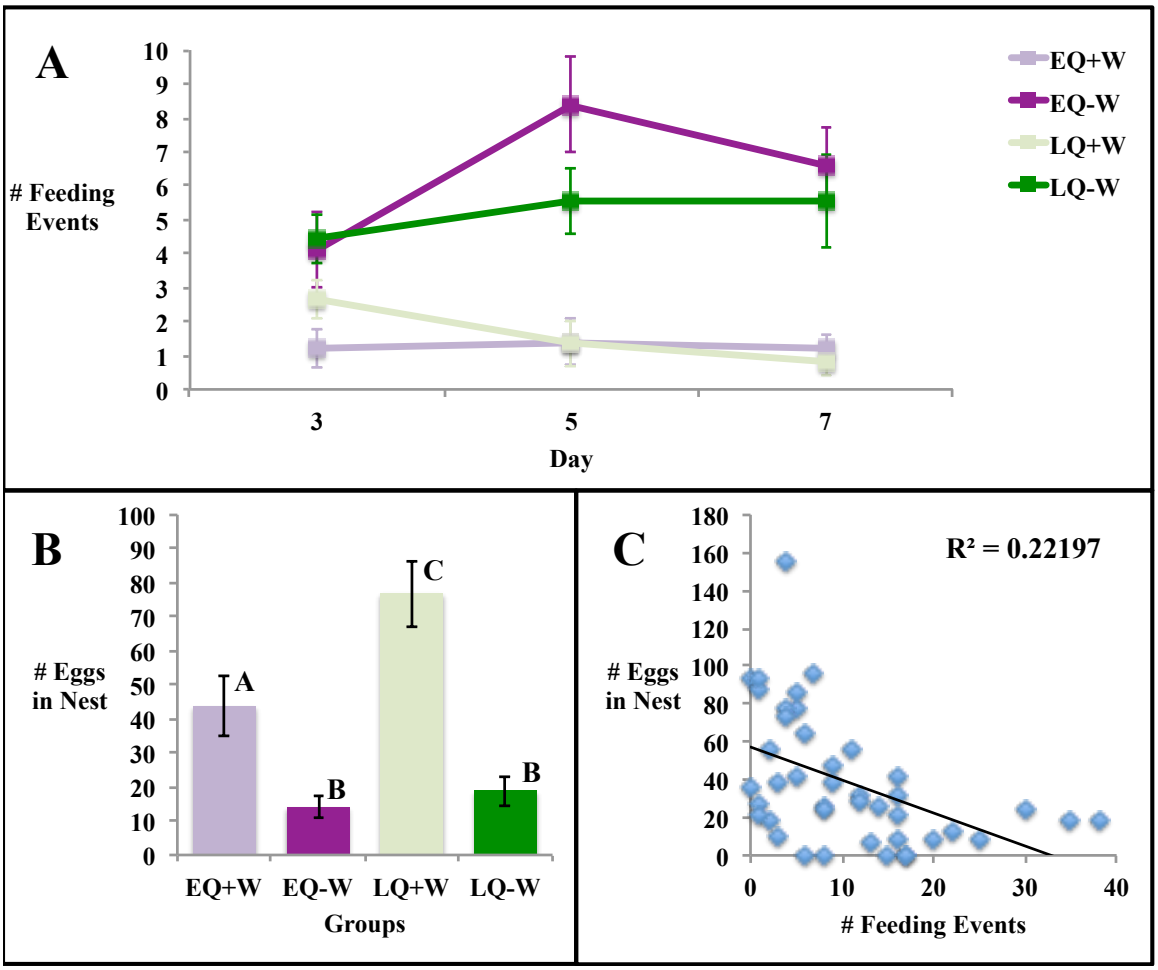


Figure 6. Effect of social manipulation on colony development. A) Effect of treatment on number of larvae in nests; *x*-axis = group, *y*-axis = number of larvae in nests on day 7. B) Effect of treatment on number of pupae in nests; *x*-axis = group, *y*-axis = number of pupae in nests on day 7. Values shown in A & B are group means with standard error; letters represent results of pairwise t-tests.

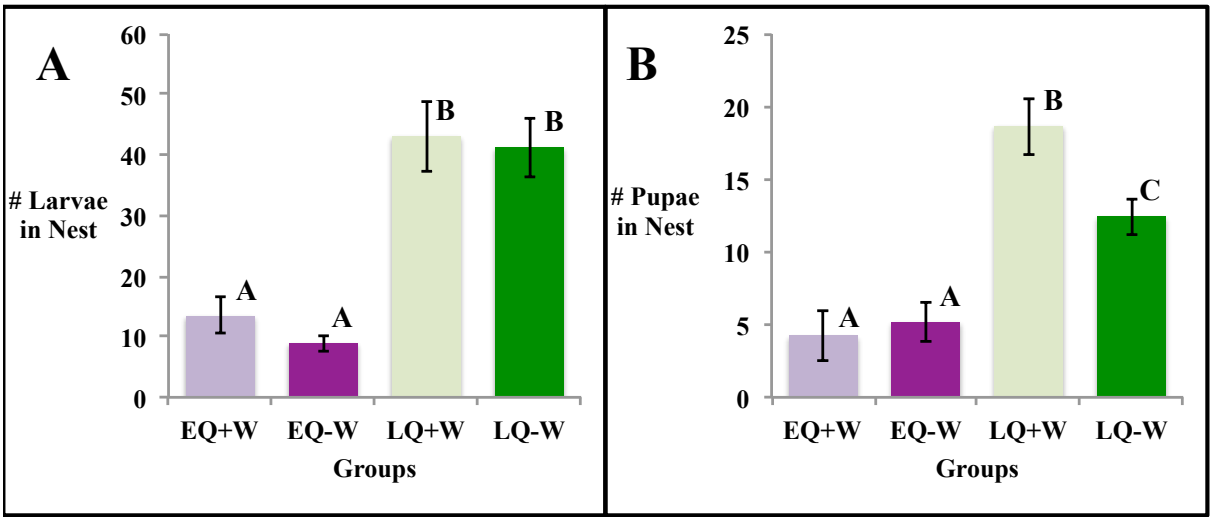


Figure 7. Comparison of gene lists from Pairwise and Rank Product Tests. "Yes" = gene(s) present in list; "No" = gene(s) not present in list.

Up in Early-Stage Queens with Workers (*vs*. Without)		**Rank Product Test** (n = 8)	
		Yes	No
Pairwise Test (n = 0)	Yes	0	0
	No	8	

Up in Queens with Workers, Early-Stage (*vs*. Late)		**Rank Product Test** (n = 28)	
		Yes	No
Pairwise Test (n = 0)	Yes	0	0
	No	28	

Down in Early Stage Queens with Workers (*vs*. Without)		**Rank Product Test** (n = 0)	
		Yes	No
Pairwise Test (n = 0)	Yes	0	0
	No	0	

Down in Queens with Workers, Early-Stage (*vs*. Late)		**Rank Product Test** (n = 18)	
		Yes	No
Pairwise Test (n = 0)	Yes	0	0
	No	18	

Up in Late-Stage Queens with Workers (*vs*. Without)		**Rank Product Test** (n = 30)	
		Yes	No
Pairwise Test (n = 303)	Yes	30	0
	No	273	

Up in Queens without Workers, Early Stage (*vs*. Late)		**Rank Product Test** (n = 453)	
		Yes	No
Pairwise Test (n = 464)	Yes	373	91
	No	80	

Down in Late-Stage Queens with Workers (*vs*. Without)		**Rank Product Test** (n = 132)	
		Yes	No
Pairwise Test (n = 7)	Yes	7	0
	No	125	

Down in Queens without Workers, Early Stage (*vs*. Late)		**Rank Product Test** (n = 566)	
		Yes	No
Pairwise Test (n = 441)	Yes	407	34
	No	159	

Table 3. Comparison of list sizes from ANOVA and Pairwise and Rank Product Tests of differential brain gene expression. "List" = gene list name; "Oligos" = number of oligos significant for test; "Unique *A. mellifera* Orthologs" = number of nonredundant *A. mellifera* orthologs, derived from oligo lists; "Unique *D. melanogaster* Orthologs" = number of nonredundant *D. melanogaster* orthologs, derived from *A. mellifera* ortholog lists. All oligos are significant at FDR-corrected $P < 0.05$.

	LIST	OLIGOS	UNIQUE *A. mellifera* ORTHOLOGS	UNIQUE *D. melanogaster* ORTHOLOGS
ANOVA	Main Effect of Workers	4	4	4
	Main Effect of Stage of Queen	394	328	253
Pairwise Tests	LQ+W *vs*. EQ+W	0	0	0
	LQ-W *vs*. EQ-W	1101	884	824
	EQ+W *vs.* EQ-W	0	0	0
	LQ+W *vs.* LQ-W	37	37	34
Rank Product Tests	Upregulated in LQ+W *vs*. EQ+W	39	28	26
	Downregulated in LQ+W *vs*. EQ+W	21	18	18
	Upregulated in LQ-W *vs*. EQ-W	575	453	430
	Downregulated in LQ-W *vs*. EQ-W	685	566	528
	Upregulated in EQ+W *vs*. EQ-W	9	8	8
	Downregulated in EQ+W *vs*. EQ-W	0	0	0
	Upregulated in LQ+W *vs*. LQ-W	338	304	286
	Downregulated in LQ+W *vs*. LQ-W	149	132	123

Table 4. Functional terms enriched in the lists of differentially expressed genes described in Table 1 (ANOVA). GO = Gene Ontology; Enrich. = fold enrichment. Gene lists used for enrichment analysis significant at FDR-corrected $P < 0.05$. Terms with < 5 genes are not included. List size for main effect of workers was too small for enrichment analysis.

	TERM	GO ID	ENRICH.	*P* - VALUE
	Main Effect of Stage			
Biological Processes	chitin metabolic process	GO:0006030	5.55	0.00891
	photoreceptor cell differentiation	GO:0046530	2.56	0.02020
	chromosome segregation	GO:0007059	2.71	0.02326
	cellular ion homeostasis	GO:0006873	4.07	0.02848
	cognition	GO:0050890	2.13	0.02849
	compound eye photoreceptor cell differentiation	GO:0001751	2.57	0.03063
	negative regulation of nitrogen compound metabolic process	GO:0051172	2.18	0.03432
	negative regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	GO:0045934	2.18	0.03432
	behavior	GO:0007610	1.84	0.03524
	ion homeostasis	GO:0050801	3.82	0.03561
	eye photoreceptor cell differentiation	GO:0001754	2.44	0.03943
	negative regulation of RNA metabolic process	GO:0051253	2.24	0.04133
	cellular chemical homeostasis	GO:0055082	3.59	0.04370
	neurological system process	GO:0050877	1.59	0.04536
	negative regulation of transcription	GO:0016481	2.20	0.04590
Molecular Functions	chitin binding	GO:0008061	6.28	0.00515
	solute:cation symporter activity	GO:0015294	3.23	0.03180
	solute:sodium symporter activity	GO:0015370	3.77	0.03665
	adenyl nucleotide binding	GO:0030554	1.32	0.04459
	symporter activity	GO:0015293	2.95	0.04556
	polysaccharide binding	GO:0030247	3.54	0.04559
	pattern binding	GO:0001871	3.54	0.04559
	purine nucleoside binding	GO:0001883	1.32	0.04827

Table 5. Functional terms enriched in the lists of differentially expressed genes described in Table 1 (Pairwise Tests). GO = Gene Ontology; Enrich. = fold enrichment. Gene lists used for enrichment analysis significant at FDR-corrected $P < 0.05$. Terms with < 5 genes are not included. The following lists were too small to perform enrichment analysis: LQ+W *vs*. EQ+W, EQ+W *vs.* EQ-W, LQ+W *vs.* LQ-W.

	TERM	GO ID	ENRICH.	*P* - VALUE
	Upregulated in LQ-W vs. EQ-W			
Biological Processes	chromatin organization	GO:0006325	1.93	0.01395
	chromosome organization	GO:0051276	1.71	0.01555
	translation	GO:0006412	1.53	0.01584
	DNA packaging	GO:0006323	2.71	0.02148
	regulation of Rab GTPase activity	GO:0032313	3.31	0.02709
	regulation of Rab protein signal transduction	GO:0032483	3.31	0.02709
	cellular macromolecular complex subunit organization	GO:0034621	1.67	0.03372
	regulation of hydrolase activity	GO:0051336	2.27	0.03698
	regulation of Ras GTPase activity	GO:0032318	2.68	0.03769
	nucleosome organization	GO:0034728	2.68	0.03769
	cellular cation homeostasis	GO:0030003	3.67	0.03781
	chromosome segregation	GO:0007059	2.20	0.04307
	regulation of Ras protein signal transduction	GO:0046578	2.15	0.04980
Molecular Functions	metal ion binding	GO:0046872	1.34	0.00128
	ion binding	GO:0043167	1.32	0.00213
	cation binding	GO:0043169	1.31	0.00264
	adenyl nucleotide binding	GO:0030554	1.41	0.00412
	transition metal ion binding	GO:0046914	1.35	0.00461
	purine nucleoside binding	GO:0001883	1.40	0.00470
	nucleoside binding	GO:0001882	1.39	0.00572
	ATP binding	GO:0005524	1.40	0.00620
	adenyl ribonucleotide binding	GO:0032559	1.39	0.00662
	purine nucleotide binding	GO:0017076	1.29	0.01472
	purine ribonucleotide binding	GO:0032555	1.28	0.02010
	ribonucleotide binding	GO:0032553	1.28	0.02010
	zinc ion binding	GO:0008270	1.31	0.02782
	nucleotide binding	GO:0000166	1.23	0.02829
	aminopeptidase activity	GO:0004177	3.83	0.03206
	DNA binding	GO:0003677	1.35	0.04143
	Downregulated in LQ-W vs. EQ-W			
Biological Processes	ion transport	GO:0006811	2.06	0.00137
	anion transport	GO:0006820	3.12	0.03512
	carbohydrate binding	GO:0030246	2.88	0.00903
	anion transmembrane transporter activity	GO:0008509	2.70	0.01374
	transition metal ion transmembrane transporter activity	GO:0046915	4.64	0.01609
	polysaccharide binding	GO:0030247	3.48	0.02207

Table 5 (continued)

	pattern binding	GO:0001871	3.48	0.02207
	di-, tri-valent inorganic cation transmembrane transporter activity	GO:0015082	4.22	0.02319
	symporter activity	GO:0015293	2.83	0.03009
	structural molecule activity	GO:0005198	1.52	0.03670

Table 6. Functional terms enriched in the lists of differentially expressed genes described in Table 1 (Rank Product tests). GO = Gene Ontology; Enrich. = fold enrichment. Gene lists used for enrichment analysis significant at FDR-corrected $p < 0.05$. Terms with < 5 genes are not included. The following lists were too small to perform enrichment analysis: Upregulated in LQ+W *vs*. EQ+W, Downregulated in LQ+W *vs*. EQ+W, Upregulated in EQ+W *vs*. EQ-W, Downregulated in EQ+W *vs*. EQ-W.

	TERM	GO ID	ENRICH.	*P* - VALUE
	Upregulated in LQ+W vs. LQ-W			
Biological Processes	ion transport	GO:0006811	2.35	0.00040
	photoreceptor cell differentiation	GO:0046530	2.91	0.00356
	regulation of Ras protein signal transduction	GO:0046578	3.30	0.00467
	adult behavior	GO:0030534	3.30	0.00467
	regulation of Rho protein signal transduction	GO:0035023	6.21	0.00644
	regulation of small GTPase mediated signal transduction	GO:0051056	2.90	0.01029
	photoreceptor cell fate commitment	GO:0046552	4.03	0.01372
	sodium ion transport	GO:0006814	5.04	0.01417
	neuron fate commitment	GO:0048663	3.87	0.01631
	cation transport	GO:0006812	2.08	0.01982
	tube development	GO:0035295	2.69	0.02585
	regulation of cell morphogenesis	GO:0022604	2.31	0.02620
	regulation of cell shape	GO:0008360	2.46	0.02649
	adult locomotory behavior	GO:0008344	3.23	0.03412
	neuron differentiation	GO:0030182	1.63	0.03809
	compound eye photoreceptor fate commitment	GO:0001752	3.67	0.04267
	alcohol catabolic process	GO:0046164	3.67	0.04267
	eye photoreceptor cell fate commitment	GO:0042706	3.67	0.04267
	metal ion transport	GO:0030001	2.23	0.04408
	respiratory system development	GO:0060541	2.10	0.04495
	open tracheal system development	GO:0007424	2.10	0.04495
	tube morphogenesis	GO:0035239	2.63	0.04591
	cellular carbohydrate catabolic process	GO:0044275	3.51	0.04925
	monovalent inorganic cation transport	GO:0015672	2.35	0.04969
	compound eye photoreceptor cell differentiation	GO:0001751	2.35	0.04969
Molecular Functions	hexose transmembrane transporter activity	GO:0015149	7.81	0.00248
	glucose transmembrane transporter activity	GO:0005355	7.81	0.00248
	monosaccharide transmembrane transporter activity	GO:0015145	7.10	0.00370
	Ras guanyl-nucleotide exchange factor activity	GO:0005088	6.01	0.00725
	cofactor binding	GO:0048037	2.23	0.00838
	NAD or NADH binding	GO:0051287	4.46	0.00879
	sugar transmembrane transporter activity	GO:0051119	5.58	0.00966
	small GTPase regulator activity	GO:0005083	2.52	0.01536
	anion transmembrane transporter activity	GO:0008509	2.66	0.02739

Table 6 (continued)

	GTPase regulator activity	GO:0030695	2.05	0.03882
	coenzyme binding	GO:0050662	2.14	0.04020
	nucleoside-triphosphatase regulator activity	GO:0060589	1.97	0.04769
	Downregulated in LQ+W vs. LQ-W			
Biological Processes	translation	GO:0006412	2.20	0.00847
	cell cycle phase	GO:0022403	2.21	0.01633
	mitotic spindle elongation	GO:0000022	3.91	0.01668
	cell cycle process	GO:0022402	2.10	0.01705
	spindle elongation	GO:0051231	3.84	0.01790
	M phase	GO:0000279	2.10	0.03146
	cell cycle	GO:0007049	1.82	0.04596
Molecular Functions	structural constituent of ribosome	GO:0003735	2.74	0.00910
	structural molecule activity	GO:0005198	2.01	0.04095
	Upregulated in LQ-W vs. EQ-W			
Biological Processes	chromosome organization	GO:0051276	1.98	0.00096
	chromatin organization	GO:0006325	2.21	0.00166
	chromosome segregation	GO:0007059	2.70	0.00344
	M phase	GO:0000279	1.70	0.00414
	cell cycle phase	GO:0022403	1.64	0.00698
	spindle organization	GO:0007051	1.80	0.00831
	mitotic spindle elongation	GO:0000022	2.31	0.01165
	dsRNA transport	GO:0033227	4.05	0.01204
	chromosome condensation	GO:0030261	4.05	0.01204
	cell cycle	GO:0007049	1.49	0.01266
	chromatin modification	GO:0016568	2.10	0.01291
	spindle elongation	GO:0051231	2.27	0.01328
	regulation of hydrolase activity	GO:0051336	2.51	0.01450
	mitotic cell cycle	GO:0000278	1.58	0.01635
	mitotic spindle organization	GO:0007052	1.77	0.01799
	microtubule cytoskeleton organization	GO:0000226	1.60	0.01957
	regulation of Rab protein signal transduction	GO:0032483	3.60	0.02027
	regulation of Rab GTPase activity	GO:0032313	3.60	0.02027
	programmed cell death	GO:0012501	2.03	0.02199
	microtubule-based process	GO:0007017	1.50	0.02228
	regulation of Ras GTPase activity	GO:0032318	3.02	0.02303
	cell cycle process	GO:0022402	1.48	0.02339
	DNA packaging	GO:0006323	2.70	0.02373
	death	GO:0016265	2.00	0.02442
	cell death	GO:0008219	2.00	0.02442
	apoptosis	GO:0006915	2.62	0.02782
	chromatin remodeling	GO:0006338	2.62	0.02782
	cellular amino acid biosynthetic process	GO:0008652	3.24	0.03149
	regulation of Ras protein signal transduction	GO:0046578	2.21	0.04490
	regulation of GTPase activity	GO:0043087	2.61	0.04527

Table 6 (continued)

Molecular Functions	ATP binding	GO:0005524	1.47	0.00110
	adenyl ribonucleotide binding	GO:0032559	1.47	0.00116
	ribonucleotide binding	GO:0032553	1.41	0.00125
	purine ribonucleotide binding	GO:0032555	1.41	0.00125
	adenyl nucleotide binding	GO:0030554	1.44	0.00131
	purine nucleoside binding	GO:0001883	1.44	0.00146
	purine nucleotide binding	GO:0017076	1.38	0.00166
	nucleoside binding	GO:0001882	1.43	0.00173
	nucleotide binding	GO:0000166	1.33	0.00216
	exopeptidase activity	GO:0008238	2.61	0.02799
	Rab GTPase activator activity	GO:0005097	3.13	0.03578
	aminopeptidase activity	GO:0004177	3.73	0.03768
	metal ion binding	GO:0046872	1.20	0.04242
	Downregulated in LQ-W vs. EQ-W			
Biological Processes	ion transport	GO:0006811	2.01	0.00010
	sodium ion transport	GO:0006814	3.84	0.00608
	long-term memory	GO:0007616	4.39	0.00737
	positive regulation of macromolecule biosynthetic process	GO:0010557	2.20	0.00824
	positive regulation of transcription	GO:0045941	2.24	0.00982
	positive regulation of cellular biosynthetic process	GO:0031328	2.06	0.01069
	positive regulation of biosynthetic process	GO:0009891	2.06	0.01069
	positive regulation of gene expression	GO:0010628	2.20	0.01150
	positive regulation of nitrogen compound metabolic process	GO:0051173	2.15	0.01339
	positive regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	GO:0045935	2.15	0.01339
	photoreceptor cell fate commitment	GO:0046552	2.93	0.01424
	cation transport	GO:0006812	1.74	0.01681
	neuron fate commitment	GO:0048663	2.81	0.01790
	memory	GO:0007613	3.07	0.01988
	positive regulation of macromolecule metabolic process	GO:0010604	1.89	0.02792
	biological adhesion	GO:0022610	1.78	0.02942
	cell adhesion	GO:0007155	1.78	0.02942
	compound eye photoreceptor fate commitment	GO:0001752	2.80	0.03150
	eye photoreceptor cell fate commitment	GO:0042706	2.80	0.03150
	monovalent inorganic cation transport	GO:0015672	1.92	0.04138
	wing disc development	GO:0035220	1.55	0.04394
Molecular Functions	L-amino acid transmembrane transporter activity	GO:0015179	5.35	0.00867
	symporter activity	GO:0015293	2.59	0.01080
	solute:sodium symporter activity	GO:0015370	3.15	0.01739
	hexose transmembrane transporter activity	GO:0015149	4.28	0.02151

Table 6 (continued)

glucose transmembrane transporter activity	GO:0005355	4.28	0.02151
solute:cation symporter activity	GO:0015294	2.48	0.02222
alkali metal ion binding	GO:0031420	3.42	0.02334
structural molecule activity	GO:0005198	1.44	0.02490
vitamin B6 binding	GO:0070279	2.85	0.02845
pyridoxal phosphate binding	GO:0030170	2.85	0.02845
monosaccharide transmembrane transporter activity	GO:0015145	3.89	0.03075
anion transmembrane transporter activity	GO:0008509	2.00	0.04062

Table 7. List of genes with expression values correlated with maternal traits. "*A. mellifera* Gene ID" = *A. mellifera* orthologs derived from oligos. *P*-values reported are for Pearson correlation coefficients.

***A. mellifera* Gene ID**	**Gene Name**	***P* - value**
	Expression correlated with brood feeding behavior	
GB19602	*sodium-coupled neutral amino acid transporter 7*	0.00020
GB13198	hypothetical protein LOC411299	0.00037
GB12627	hypothetical LOC551126	0.00039
GB14934	*tryptophanyl-tRNA synthetase*	0.00042
GB16129	*succinate dehydrogenase cytochrome b small subunit, mitochondrial*	0.00044
GB16868	*feminizer*	0.00047
GB17688	*inosine triphosphate pyrophosphatase*	0.00050
GB17957	*ras-like protein 2*	0.00058
GB18234	*high affinity copper uptake protein 1*	0.00058
GB13850	*cleavage and polyadenylation specificity factor 100*	0.00061
GB16412	*translationally controlled tumor protein*	0.00063
GB17536	unknown	0.00066
GB19751	*alpha mannosidase II*	0.00068
GB18025	*beta-catenin-like protein 1*	0.00085
GB15724	*zinc finger C4H2 domain-containing protein*	0.00087
GB18632	*uncharacterized MFS-type transporter C09D4.1*	0.00088
GB12825	*cleavage and polyadenylation factor subunit*	0.00092
	Expression correlated with egg laying behavior	
GB10161	*venom acid phosphatase Acph*	0.00018
GB18510	*threonine aspartase 1*	0.00024
GB13341	*THO complex subunit 6*	0.00033
GB16201	hypothetical LOC408320	0.00065
GB13722	*glucocerebrosidase*	0.00080
GB12816	*teneurin-3*	0.00084
GB13833	*thrombospondin*	0.00088
GB10483	*venom acid phosphatase Acph-1*	0.00090

LITERATURE CITED

1. Wilson EO (2000) *Sociobiology: The New Synthesis.* Belknap Press, Harvard.
2. Free JB, Butler CG (1959) *Bumblebees*. Collins, London.
3. White SA, Nguyen T, Fernald RD (2002) Social regulation of gonadotropin-releasing hormone. *J Exp Biol* 17:2567-2581.
4. Ratnieks FLW, Foster KR, Wenseleers T (2006) Conflict resolution in insect societies. *Ann Rev Entomol* 51: 581-608.
5. Robinson GE (1992) Regulation of division of labor in insect societies. *Ann Rev Entomol* 37:637-665.
6. Eban-Rothschild A, Belluci S, Bloch G (2011) Maternity-related plasticity in circadian rhythms of bumble-bee queens. *Proc Biol Sci, In Press.*
7. Dornhaus A, Brockmann A, Chittka L (2003) Bumble bees alert to food with pheromone from tergal gland. *J Comp Physiol A* 189:47-51.
8. Alaux C, Jaisson P, Hefetz A (2004) Queen influence on worker reproduction in bumblebees (*Bombus terrestris*) colonies. *Insectes Soc* 51:287-293.
9. Bloch G, Hefetz A (1999) Regulation of reproduction by dominant workers in bumblebee (*Bombus terrestris*) queenright colonies. *Behav Ecol Sociobiol* 45:125-135.
10. Gretenkord C, Drescher W (1997) Successful colony foundation and development of experimentally hibernated *Bombus terrestris* queens depending on different starting methods. *Acta Hort* 437:271–276.
11. Gurel F, Gosterit A (2008) Effects of different stimulation methods on colony initiation and development of *Bombus terrestris* L.(Hymenoptera: Apidae) queens. *App Entomol Zool* 43:113-117.
12. Kwon YJ, Than KK, Suh SJ (2006) New method to stimulate the onset of *Bombus terrestris* (Hymenoptera: Apidae) rearing: Using worker helpers in the presence of frozen pupae. *Entomol Res* 36:202-207.

13. Robinson GE, Fernald RD, Clayton DF (2009) Genes and social behavior. *Science* 322:896-900.
14. Woodard SH, *et al*. (2010) Genes involved in convergent evolution of eusociality in bees. *Proc Natl Acad Sci USA* 108:7472-7477.
15. Robinson GE (2000) Sociogenomics: Social life in molecular terms. *Nat Rev Genet* 6:257-271.
16. Geva S, Hartfelder K, Bloch G (2005) Reproductive division of labor, dominance, and ecdysteroid levels in hemolymph and ovary of the bumble bee *Bombus terrestris*. *J Insect Physiol* 51:811-823.
17. Yerushalmi S, Bodenhaimer S, Bloch G (2006) Developmentally determined attenuation in circadian rhythms links chronobiology to social organization in bees. *J Exp Biol* 209:1044-1051.
18. Velthuis HHW, Van Doorn A (2006) A century of advances in bumblebee domestication and the economic and environmental aspects of its commercialization for pollination. *Apidologie* 37:421-451.
19. Free JB, Weinberg I, Whiten A (1969) The egg-eating behaviour of *Bombus lapidarius* L. *Behaviour* 35:313–317.
20. Duchateau MJ (1989) Agonistic behaviour in colonies of the bumblebee *Bombus terrestris*. *J Ethology* 7:141–151.
21. Schulz DJ, Robinson GE (1999) Biogenic amines and division of labor in honey bee colonies: Behaviorally related changes in the antennal lobes and age-related changes in the mushroom bodies. *J Comp Physiol A* 184:481-488.
22. Alaux C, *et al*. (2009). Honey bee aggression supports a link between gene regulation and behavioral evolution. 106:15400-15405.
23. Sen Sarma M, Rodriguez-Zas S, Hong F, Zhong S, Robinson GE (2009) Transcriptomic profiling of central nervous system regions in three species of honey bee during dance communication behavior. *PLoS One* 4:e6408.

24. R Development Core Team (2011) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.

25. Smyth GK (2005) Limma: linear models for microarray data. In: *Bioinformatics and Computational Biology Solutions using R and Bioconductor*, Gentleman R, Carey V, Dudoit S, Irizarry R, Huber W, Eds. Springer, New York. pp 397-420.

26. Smyth GK and Speed TP (2003) Normalization of cDNA microarray data. *Methods* 31:265-273.

27. HBGSC (2006) Insights into social insects from the genome of the honeybee *Apis mellifera*. *Nature* 443:931-949.

28. Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J Roy Stat Soc B* 57:289-300.

29. Breitling R, Armengauda P, Amtmanna A, Herzyk P (2004) Rank products: a simple, yet powerful, new method to detect differentially regulated genes in replicated microarray experiments FEBS Letters 573:83–92.

30. Hong F, Breitling R, McEntee1 CW, Wittner BS (2006) RankProd: a bioconductor package for detecting differentially expressed genes in meta-analysis. *Bioinf App Note* 22:2825-2827.

31. Huang DW, Sherman BT, Lempicki RA (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protoc* 4:44-57.

32. Sasagawa H, Narita R, Kitagawa Y, Kadowaki T (2003) The expression of genes encoding visual components is regulated by a circadian clock, light environment and age in the honeybee (*Apis mellifera*). *Eur J Neurosci* 17:963-970.

33. Merrill CE, Sherertz TM, Walker WB, Zwiebel LJ (2005) Odorant-specific requirements for *arrestin* function in Drosophila olfaction. *J Neurobiol* 63:15-28.

34. Waddell S, Quinn WG (2001) Review. Flies, genes, and learning. *Ann Rev Neurosci* 24:1283-1309.

35. Agrawal N, Venkiteswaran G, Sadaf S, Padmanabhan N, Banerjee S, Hasan G (2010) Inositol 1,4,5-trisphosphate receptor and dSTIM function in *Drosophila* insulin-producing neurons regulates systemic intracellular calcium homeostasis and flight. *J Neuroscience* 30:1301-1313.

36. Banerjee S, Lee J, Venkatesh K, Wu CF, Hasan G (2004) Loss of flight and associated neuronal rhythmicity in inositol 1,4,5-trisphosphate receptor mutants of *Drosophila*. *J Neuroscience* 24:7869-7878.

37. Kucharski R, Maleszka R (2002) Molecular profiling of behavioural development: differential expression of mRNAs for inositol 1,4,5-trisphosphate 3-kinase isoforms in naive and experienced honeybees (*Apis mellifera*). *Mol Brain Res* 99:92-101.

38. Whitfield CW, Cziko AM, Robinson GE (2003) Gene expression profiles in the brain predict behaviour in individual honey bees. *Science* 302:296–299.

39. Toth AL, Varala K, Henshaw MT, Rodriguez-Zas SL, Hudson ME, Robinson GE (2010). Brain transcriptomic analysis in paper wasps identifies genes associated with behaviour across social insect lineages. *Proc Roy Soc Lond B* 277:2139-2148.

40. Hung HC, Maurer C, Kay SA, Weber F (2007) Circadian transcription depends on limiting amounts of the transcription co-activator nejire/CBP. *J Biol Chem* 282:31349-31357.

41. Shemesh Y, Cohen M, Bloch G (2007) Natural plasticity in circadian rhythms is mediated by reorganization in the molecular clockwork in honeybees. FASEB J 10:2304-11.

42. Ament SA, Corona M, Pollock HS, Robinson GE (2008) Insulin signaling is involved in the regulation of worker division of labor in honey bee colonies. *Proc Natl Acad Sci USA* 105:4226-4231.

43. Page RE, Amdam GV (2007) The making of a social insect: developmental architectures of social design. *BioEssays* 29: 334–343.

44. Corona M, Velarde RA, Remolina S, Moran-Lauter A, Wang Y, Hughes KA, Robinson GE (2007) Vitellogenin, juvenile hormone, insulin signaling, and queen honey bee longevity. *Proc Natl Acad Sci USA* 104:7128-7133.

45. Ben-Shahar Y, Robichon A, Sokolowski MB, Robinson GE (2002) Influence of gene action across different time scales on behavior. Science 296:741-744.

46. Alaux C, Jaisson P, Hefetz A (2005) Reproductive decision-making in semelparous colonies of the bumblebee *Bombus terrestris*. Behav Ecol Sociobiol 59:270-277.

47. Alaux C, Le Conte Y, Adams HA, Rodriguez-Zas S, Grozinger CM, Sinha S, Robinson GE (2009) Regulation of brain gene expression in honey bees by brood pheromone. *Genes Brain Behav* 8:309-319.

48. Grozinger CM, Sharabash NM, Whitfield CW, Robinson GE (2003) Pheromone-mediated gene expression in the honey bee brain. *Proc Natl Acad Sci USA* 100:14519-14525.

49. Cummings ME, *et al*. (2008) Sexual and social stimuli elicit rapid and contrasting genomic responses. *Proc. Roy. Soc. Lond. B* 275: 393-402.

50. Replogle K, *et al*. (2008) The songbird neurogenomics (SoNG) initiative: community-based tools and strategies for study of brain function and evolution. *BMC Genomics* 9:1-20.

51. Burmeister SS, Jarvis ED, Fernald RD (2005) Rapid behavioral and genomic responses to social opportunity. *PloS Biol* 3:e363.

52. Trainor BC, Hofmann HA (2007) Somatostatin and somatostatin receptor gene expression in dominant and subordinate males of an African cichlid fish. *Behav Brain Res* 179:314-320.

53. Sabatini MJ, Ebert P, Lewis DA, Levitt P, Cameron JL, Mirnics K (2007) Amygdala gene expression correlates of social behavior in monkeys experiencing maternal separation. *J Neuroscience* 27:3295-3304.

54. Cameron SA, *et al.* (2011) Patterns of widespread decline in North American bumble bees. *Proc Natl Acad Sci USA* 108:662-667.

55. Muller CB, Schmid-Hempel P (1992) Correlates of reproductive success among field colonies of *Bombus lucorum* - the importance of growth and parasites. *Ecol Entomol* 17:343-353.

CHAPTER 5: THE MOLECULAR BASIS AND EVOLUTION OF COOPERATIVE BROOD CARE IN BUMBLE BEES

Abstract

Sibling care is one of the hallmarks of the social insects, but its evolution remains a challenge to explain. The hypothesis that sibling care evolved from ancestral maternal care has been elaborated in molecular terms to involve heterochronic changes in gene expression; this has led to the prediction that workers in primitively eusocial species will show patterns of gene expression more similar to foundress queens engaged in both maternal care and reproductive behavior than to established queens engaged solely in reproductive behavior. I tested this idea in the bumble bee *Bombus terrestris*, using a microarray platform with ca. 4,500 genes. Unlike in the wasp *Polistes metricus*, in which support for the above prediction has been obtained, I found that brain gene expression patterns in foundress and queen bumble bees were more similar to each other than to workers. However, a comparison of lists of differentially expressed genes derived from this study and gene lists from microarray studies in *Polistes* and the honey bee *Apis mellifera* suggest that there is a shared set of genes involved in the regulation of related social behaviors across independent eusocial lineages. Together, these results suggest that the multiple independent evolutions of eusociality that occurred in the insects involved a combination of shared and different mechanisms.

Introduction

One of the hallmarks of the insect societies is sibling care, where the adult female offspring of the queen, called workers, remain in their natal nest and care for their siblings. Sibling care in the social insects is multifaceted and involves tasks such as provisioning the nest, direct feeding of larvae, and cleaning of brood cells (1). Various evolutionary scenarios have been put forth to explain the evolution of sibling care in the insect societies (2), including the prominent hypothesis that sibling care evolved

from ancestral maternal care (3-5). According to this idea, reproductive (egg laying) and maternal care (brood provisioning) behavior were uncoupled temporally, eventually occurring in separate castes, queens and workers.

A molecular perspective on this hypothesis posits that sibling care originated via heterochronic changes in gene expression, such that the molecular architecture underlying maternal care behavior was co-opted to be used in a new developmental context, the worker caste (6). The hypothesis that sibling care evolved from maternal care (hereafter, the molecular heterochrony hypothesis) predicts that sibling care and maternal care behaviors are regulated by similar patterns of gene expression. This hypothesis is particularly compelling in social insect lineages where both maternal care and sibling care are present and involve strikingly similar behaviors in queens and workers. More broadly, the molecular heterochrony hypothesis is supported by studies on the evolution of development in a variety of taxa, which together suggest that evolution is conservative, and evolutionary novelty is often achieved through the co-option of pre-existing genetic architecture (7).

There also are alternatives to the hypothesis that sibling care evolved from maternal care, including the idea that sibling care evolved *de novo* (8), and therefore is not evolutionarily rooted in maternal care behavior. Support for the "*de novo* hypothesis" may be found in the complete absence of maternal care in some lineages of eusocial insects. However, studies on the honey bee *Apis mellifera*, a highly eusocial species that lacks maternal care, suggest that there may nonetheless be an evolutionary link between sibling care and maternal care in this species (9). Given that complex sociality evolved 11 or more times in the insects (10), it is possible, and perhaps even likely (11), that sibling care evolved via multiple evolutionary routes in different social insect lineages.

A related question is whether similar genes and molecular pathways have been involved in the evolution of sibling care across independent social insect lineages, regardless of the routes or evolutionary mechanisms through which sibling care evolved (7, 11). There is a growing body of

evidence that suggests that there are many commonalities in genes and molecular pathways associated with various behavioral states across the independently evolved social insect lineages, referred to as "genetic toolkits" (7). For example, the insulin signaling pathway appears to be one such genetic toolkit, as it has been implicated in the regulation of various aspects of worker division of labor in all of the major groups of social insects (the ants, bees, wasps, and termites) (12). An important goal in the study of genetic toolkits in the social insects is to determine how widespread this phenomenon is, as well as to further refine our understanding of how the same genes and pathways operate in the regulation of social behaviors in the different lineages (7).

"Primitively" eusocial insects (which include some bees and wasps) are excellent subjects to explore the evolution of sibling care because queens and workers often perform similar brood care-related tasks, albeit during different stages in the colony cycle. Prior to the emergence of the first workers in a colony, "foundress" queens must provision, feed, and otherwise tend to their offspring, tasks that are assumed by workers following their emergence in the nest. It was these features of the life history of the primitively eusocial *Polistes* wasps that inspired the hypothesis that sibling care evolved from ancestral maternal care (3-5).

Toth *et al*. (13) utilized the similarity between maternal and sibling care behavior in the primitively eusocial wasp *Polistes metricus* to test the molecular heterochrony hypothesis. Using patterns of brain gene expression for a set of 32 genes associated with worker division of labor in the honey bee *Apis mellifera*, Toth *et al*. (13) showed that foundresses and foraging workers, the two groups that forage and provision the nest, share the most similar patterns of brain gene expression relative to other females in the colony, the queens and "gynes" (non-foraging and non-reproductive individuals destined to become queens the following season). This clustering of foundresses and foraging workers based on brain gene expression data occurred despite the fact that there are major differences in reproductive status between these two groups, as *Polistes* foundresses are mated, have developed

ovaries, and lay eggs, and foraging workers are entirely non-reproductive. Using a more extensive set of genes in a microarray study, Toth *et al.* (14) showed that the above pattern was specific to the subset of genes that was tested, because for most genes in the *Polistes* genome, foundresses and queens show more similar patterns of brain gene expression to each other than to workers. These results are consistent with the hypothesis that sibling care evolved from maternal care in this lineage.

I tested the molecular heterochrony hypothesis on the evolution of sibling care in bumble bees in *Bombus terrestris,* using an Agilent microarray based on a large-scale brain transcriptomic project (11). Despite evolving eusociality independently, both *Bombus* and *Polistes* share an annual colony cycle wherein each spring, foundresses initiate new nests that will persist until the end of the summer (Fig. 8A). Following the emergence of the first offspring in the nest, foundresses are considered true queens. Most of a queen's female offspring develop into workers who perform brood care and other work-related tasks for the nest, but later in the season, some female offspring develop into gynes; they mate, overwinter, and become the foundresses (and later on, queens) of the following season.

As in Toth *et al.* (13, 14), I explored patterns of brain gene expression in foundresses, queens, gynes, and workers, to identify the relationships between these groups. Unlike Toth *et al.* (13, 14), who focused on nest provisioning behavior, which involves flight- and foraging-related activities, my experiments in *Bombus* instead focused on the direct feeding of brood (hereafter, "brood-feeding behavior") as a component of sibling and maternal care. In bumble bee colonies, developing brood are progressively fed during the larval stage, a task that involves the direct placement of food (*i.e.,* pollen and honey) in larval cells (15). As with nest provisioning, brood-feeding behavior is performed by foundresses during the nest-founding phase of the colony cycle and also later on by workers upon their emergence in the colony. Whereas *Polistes* nests openly and all individuals are exposed to light, *Bombus* are cavity nesters, and queens and workers specialized on brood-feeding behavior do not typically leave the nest. By focusing on brood-feeding behavior in *B. terrestris*, I could collect individuals that differed in

their brood care behavior but were kept under constant darkness (with the exception of a lighted foraging arena; see methods) and unable to freely forage, conditions that control for effects of light exposure and flight activity on brain gene expression (16).

I performed an additional, comparative analysis to identify genes that may have repeatedly played a role in the evolution of sociality in insects, across three independent eusocial insect lineages (17). Here, I explored the overlap between lists of genes potentially involved in reproduction and/or brood care in bumble bees (generated from this study) and lists of genes associated with brood care and reproduction in *Polistes* (13, 14) and with worker-worker division of labor in the honey bee *Apis mellifera* (16, 18). Genes found in the overlap of two or more of these gene lists, which are therefore differentially expressed in important eusocial contexts across two or more eusocial lineages, may represent components of a genetic toolkit for the evolution of complex social behaviors in the insects (7).

Methods

Bees. All collections were performed in Israel during a 2-month period in summer 2009. Bees were purchased from Polyam Industries (Kibbutz Yad-Mordechai, Israel) and reared under the conditions described in (19). The four groups (Fig. 8B) were identified and collected as follows: (i) *foundresses* (*n* = 14) were individuals in the queen caste who had recently initiated nests, had larvae in their nests, and were necessarily performing all brood care because adult workers had not yet emerged in the nests; (ii) *queens* (*n* = 16) were queens with mature nests (~50 workers) who had ceased to perform direct brood feeding; (iii) *gynes* (*n* = 11) were newly-emerged (3 d old) queens collected from 3 separate colonies; (iv) *workers* (*n* = 14) were 7-9 d old, non-reproductive workers who were specialized on feeding in the nest. To identify workers specialized on feeding brood, all workers in 3 source colonies were individually marked on the day of their emergence and the colonies were observed for a 2-week period. Each day of

observation, each colony was observed for 1 h and both the total number of brood-feeding events observed and the identity of the brood-feeding bee were recorded. Brood feeding is an easily observable behavioral sequence that lasts approximately 5-15 seconds; larvae are clumped together spatially, and brood-feeding bees go to these clumps of larvae, open and inspect larval cells, and then regurgitating food into the cells (15). When regurgitating, a bee places her mouthparts into the open cell and her abdomen contracts. Bees collected as *workers* were observed feeding larvae on 3 out of 5 observation days, including the day of or day before collection. For each colony, a pollen feeder was kept in a lighted foraging arena attached to the colony with a ~0.3 m tube. Before and after each observation period, the lighted foraging chambers were scanned and the identity of any bees in the foraging arenas was recorded; any bees observed in the lighted foraging arenas were not included in the analysis.

Ovary dissections. The ovaries of all bees collected were dissected to ensure reproductive status. The abdomens were collected and stored at -20˚C until dissections were performed using methods described in (19). For *workers*, the average lengths of the 2 largest terminal oocytes (one per ovariole) were measured using an optical scale, and these values were averaged as a measure of ovary development, as in (20). All *workers* in the study had values < 1.2 mm. For all other groups (*foundresses, queens*, and *gynes*), ovaries were classified as developed or undeveloped based on gross morphology. As has previously been reported (21), there was little variation in ovary development within these three groups; all *foundresses* and *queens* had fully developed ovaries, and all *gynes* had completely undeveloped ovaries.

RNA preparation. All bees were collected directly into dry ice and their heads were immediately removed, placed in liquid Nitrogen, and stored at -80˚C to ensure RNA preservation. For brain

dissections, whole bee heads were partially lyophilized and the dissections were performed over dry ice (22). RNA was isolated from dissected brains using an RNeasy Mini Kit (QIAGEN), following the kit protocol except that the initial homogenization was done in a 500 ul microfuge tube using 100 ul extraction buffer.

Microarrays. Agilent 4 x 44K *B. terrestris* Brain EST-based microarrays were designed from sequence data obtained from 454/Roche pyrosequencing of *B. terrestris* mRNA derived primarily from brain tissue (90% and 10% abdomen) (11). Each microarray slide contained 4 identical arrays, each containing approximately 44K EST-based probes corresponding to a total of 4,503 putative *A. mellifera* genes (annotation methods described below). 250-1000 ng total RNA per bee sample was reverse transcribed and linear amplified according to manufacturer's instructions (Agilent Technologies). Samples were hybridized on the microarray slide and washed according to the Agilent protocol. Slides were scanned using an Axon 4000B scanner and images were analyzed with GENEPIX software (Agilent Technologies, Santa Clara, CA). Methods are described in greater detail in (18, 23).

Statistical analysis of microarray data. Microarray data pre-processing and statistical analyses were done in R (24) using the limma package (25). Median foreground and median background values from the 55 .gpr files from all 4 groups were read into R and any spots that had been manually flagged (-100 values) were given a weight of zero (25). The background values were ignored because investigations showed that trying to use them to adjust for background fluorescence added more noise to the data; background was low and even for all microarrays, so no background correction was done. Because only one sample was hybridized to each microarray slide, the single-channel expression values were normalized with the quantile method (26) then log2 transformed.

Coefficient of variation (CV) values were calculated using the 55 samples; the distribution of CV values was strongly bimodal, separating at a value of 0.015 (data not shown), suggesting a subset of probes with little variation overall and hence no useful information; nearly all of the positive and negative control spots had low CV values as well. All control spots were then removed, as well as probes with CV values < 0.015, leaving 36,876 spots out of 45,220.

A 2x2 factorial model (27) (main effects: reproduction, feeding) was fit in limma, taking into account variation due to possible differences in each microarray slide (28). The overall ANOVA F-test, the main effects of reproduction and feeding, and the interaction term were estimated as well as the mean expression level for the 4 groups. Multiple test adjustment was done separately for each contrast using the False Discovery Rate method (29).

All probes on the microarray were annotated based on orthology to the honey bee *A. mellifera* by performing BLAST to the non-redundant nucleotide database in GenBank. Probes that did not have strongly supported (E-value < 0.0001) hits to genes in the Official Honey Bee Gene Set Version 2 for *A. mellifera* (30) were excluded from further analyses, leaving 10,037 annotated probes putatively representing 4503 unique *A. mellifera* genes. Lists of genes potentially associated with brood-feeding behavior and with reproduction (the "Feeding Gene List" and Reproduction Gene List", respectively) were created from the lists of annotated probes that were significant at an FDR-corrected $p < 0.05$ for each of these main effects in the ANOVA. When multiple probes corresponding to the same *A. mellifera* gene were significant for a main effect but regulated in opposite directions, the corresponding genes were removed from the Feeding and Reproduction Gene Lists.

To identify gene functional terms enriched in the Feeding and Reproduction Gene Lists, these lists were transformed into lists of *Drosophila melanogaster* orthologs using a previously published *A. mellifera - D. melanogaster* ortholog list (30), and analyzed using the Gene Ontology (GO) functional annotation tool (GOFat level) on the DAVID website (31). Genes present in both the Feeding and

Reproduction Gene Lists, as well as genes significant for the interaction term in the ANOVA, were removed from the lists prior to identifying *D. melanogaster* orthologs; these filtered genes were combined to create a third list (the "Both Gene List"). For each of the three gene lists, enriched ($p < 0.05$; > 5 genes per term) *biological process* and *molecular function* terms were identified using a background list consisting of 3686 *D. melanogaster* genes corresponding to the 4503 putative *A. mellifera* genes on the microarray.

Patterns of gene expression across four groups. To explore patterns of brain gene expression across the four groups, Hierarchical Clustering Analysis (HCA) was performed on multiple sets of probes using the stats (24) and gplots (32) R packages. First, to identify overall patterns, HCA was performed on the 4 mean group estimates for all 8044 probes significant for the F test. Next, to best approximate the analyses used by Toth *et al.* (13, 14), HCA was also performed on two subsets of probes: 1) The 45 *B. terrestris* probes corresponding to the 32 feeding- and reproduction-related genes originally used to measure patterns of brain gene expression in *Polistes* (13) ("Test 1"); and 2) A broader subset of 1022 probes corresponding to genes implicated in reproduction and/or feeding in insects in other studies ("Test 2"). The probes used for Test 2 were those on the microarray corresponding to genes that (*i*) have feeding- and nutrition-related functions in *D. melanogaster* (*i.e.,* their *D. melanogaster* orthologs have associated *biological process* or *molecular function* Gene Ontology terms in that include the words "food", "feed", "nutrient", or "foraging"; $n = 17$); or (*ii*) are involved in reproduction-related functions in *D. melanogaster* (*i.e.,* their *D. melanogaster* orthologs have associated *biological process* or *molecular function* Gene Ontology terms that include the words "reproductive", "reproduction", "oocyte", "oogenesis", or "gamete"; $n = 222$); or (*iii*) were used in Toth *et al.* (13) ($n = 19$); in addition to (*iv*) the 27 genes differentially expressed between *A. mellifera* workers specialized on in-hive tasks versus foraging tasks ("nurses" and "foragers", respectively) in (16) and/or (18) and also differentially expressed

between provisioning and non-provisioning *Polistes* wasps in (14) and between brood-feeding and non-brood-feeding individuals in this study; and (*v*) 35 genes differentially expressed between reproductive and non-reproductive *Polistes* wasps in (14) and also between reproductive and non-reproductive bumble bees in this study. Some probes used for Test 2 met more than one of the criteria. Finally, to discern whether the patterns from Tests 1 and 2 were specific to these subsets, HCA was performed using 200 random subsets of 45 and 1022 probes and the topologies from this analysis were compared to the topologies resulting from Tests 1 and 2.

Cross-species comparisons. Both Pairwise Fisher's Exact Tests (one-tailed) and a simulation-based method (also used and described in 14) were used to determine whether a statistically significant ($p < 0.05$) degree of overlap exists between the *Bombus* Reproduction and Feeding Gene Lists and lists of genes associated with reproduction and feeding generated from a brain microarray experiment in *Polistes* (14) and lists of genes differentially expressed in *A. mellifera* nurse and forager brains, also generated from microarray experiments (16, 18). Both nurse and forager honey bees undertake feeding-related brood care tasks, with nurses directly interacting with and feeding larvae, and foragers leaving the nest to collect food provisions. Nurse and forager honey bees are not typically reproductive, although nurses may possess some "reproductive-like" phenotypic traits (9). All cross-species comparisons were performed using *A. mellifera* orthologs, and background lists for each comparison were limited to genes represented on both species' microarray. Genes were counted as overlapping regardless of the direction of change in the two lists.

Results

ANOVA of microarray data. Of the 10,037 probes included in the statistical analysis, 80% (*n* = 8044) had overall ANOVA F-tests with FDR-corrected *p*-values < 0.05, putatively representing 4040 unique genes differentially expressed across the four groups.

Feeding Gene List. 2562 genes had levels of brain expression associated with differences in brood-feeding behavior in the ANOVA. This list included the gene *insulin receptor substrate 1*, which was upregulated in brood-feeding individuals, and was also upregulated in nest-provisioning *Polistes* in Toth *et al.*'s study (13). Insulin signaling has repeatedly been implicated as playing an important role in social feeding and foraging behaviors, across several social insect lineages (12, 33). Also present in the list was the gene *kruppel homolog 1*, a gene that appears to be regulated by Queen Mandibular Pheromone (QMP) in honey bees (34) and also by both the queen and Juvenile Hormone (JH) in *B. terrestris* (35). The *biological process* and *molecular* function terms enriched in the Feeding Gene List (and in the Reproduction and Both Gene Lists) are included in Table 8.

Reproduction Gene List. 3112 genes had levels of expression levels associated with differences in reproductive status in the ANOVA. This list included another insulin signaling-related gene, *target of rapamycin* (*tor*), which was upregulated in reproductive individuals. One term related to reproduction (*gamete generation*) was enriched in the Reproduction Gene List, as well as several terms related to nervous system development (*e.g., axonogenesis*) and aging (*e.g., determination of adult life span*) (Table 8).

Both Gene List. Within the Both Gene List (*n* = 132 genes), the *biological process* terms *behavior* and *locomotory behavior* were enriched, as well as several terms related to oocyte development (*e.g., oocyte fate determination*) and the term *reproductive behavior* (Table 8).

Patterns of expression across the four groups. HCA performed using expression data from all probes significant in the overall ANOVA yielded the following relationships between the four groups: (*worker* (*gyne* (*foundress*, *queen*))) (Fig. 9A). For comparison, Figure 9 (B) also shows relationships between the same four groups in *Polistes*, which Toth *et al.* (14) derived from HCA of brain expression data using the 447 probes significant for the overall ANOVA in their *Polistes* microarray experiment.

For Test 1, HCA yielded the following relationships between the four groups: (*gyne* (*worker* (*foundress*, *queen*))) (Fig. 10A). Table 9 provides a summary of expression data for the 32 genes in Toth *et al.*'s (13) original *Polistes* study, including results from both *Polistes* studies (13, 14) and this study. For Test 2, HCA yielded the same relationships obtained from HCA of all probes significant for the overall ANOVA: (*worker* (*gyne* (*foundress*, *queen*))) (Fig. 10B). For comparison, Fig. 10 (C) also includes a heat map recreated from the *Polistes* data in ref. (13).

In 100% of the 200 random subsets of 45 and 1022 probes, *foundresses* and *queens* clustered together, with *workers* and *gynes* clustering in various ways.

Cross-species comparisons. For the tests of overlap between the *Bombus* (this study), *Polistes* (14), and *Apis* (16, 18) gene lists, only the comparison between the *Bombus* Feeding Gene List and the list of genes associated with nest provisioning in *Polistes* (14) showed a significant ($p < 0.05$ for Fisher's Exact Test; $p < 0.01$ for simulation based-test) degree of overlap (Fig. 11). Tables 10 and 11 show genes present in the overlap of two or more studies. Among the seven genes present in the brood care-related lists from all three species was the gene *inositol 1,4,5,-tris-phosphate receptor* (*itpr1*), a component of the inositol signaling pathway, which is involved in both larval feeding (36) and flight behavior (37) in

Drosophila. This gene was downregulated in in brood-feeding *Bombus*, nest-provisioning *Polistes* (14), and *Apis* foragers (18).

Discussion

The evolutionary origins of sibling care are a major topic of research on the social insects. Here, I used Agilent brain-EST based microarrays to examine patterns of brain gene expression in bumble bees (*B. terrestris*) differing in their brood care and reproductive behavior, in order to explore hypotheses on the evolution of sibling care in this lineage. I found that for most genes on the microarray, as well as for subsets of genes with functions related to feeding and reproduction, foundresses and queens share the most similar patterns of brain gene expression.

By contrast, studies on the wasp *Polistes metricus* have shown that foundress and worker share the most similar patterns of brain gene expression for a subset of 32 genes implicated in division of labor in honey bees using a quantitative PCR-based analysis (13), whereas queens and workers showed the most similar patterns for ~3000 genes analyzed by microarray (14). In both *Bombus* and *Polistes*, foundresses and queens are reproductive but they differ in their brood care behavior, suggesting that in *Bombus*, but not in *Polistes*, reproductive status has a stronger influence on overall patterns of brain gene expression than brood-feeding behavior. The strong, shared patterns of expression that I detected in foundress and queen bumble bees may also suggest that at the level of brain gene expression, fewer changes occur during the transition from foundress to queen in the *Bombus* life cycle relative to *Polistes*, although in both lineages a complex suite of physiological and behavioral changes are involved in this transition (21, 38).

In natural contexts, the transition from foundress to queen in *Bombus* may be viewed as more involved than in *Polistes*, as in *Bombus* this transition involves cessation of flight and exposure to light, whereas in *Polistes,* which nest openly, foundresses and queens are both exposed to light and queens

may occasionally fly from the nest. Additionally, foundresses in many *Polistes* species (including *P. metricus*) may found nests in small groups, in which case the dominant foundresses behaves "queen-like" (39), suggesting that there may be greater flexibility in the transition from foundress and queen in *Polistes* relative to *Bombus*.

The results of my HCA suggest that perhaps in the bumble bees, sibling care is not evolutionarily rooted in ancestral maternal care behavior, and therefore may have evolved *de novo*. Alternatively, sibling care may have evolved from maternal care in this lineage, but predictions consistent with the molecular heterochrony hypothesis were not upheld perhaps because evolutionary diversification in brain gene expression patterns in queens and workers precludes detecting similarities in brain transcription associated with sibling and maternal care.

An additional, methodological explanation for the different patterns observed in *Bombus* and in *Polistes* (13, 14) is the focus here on the direct brood-feeding component of brood care, rather than on nest provisioning, as was focused on in *Polistes*. Foraging and other flight-related behaviors have strong, documented effects on brain gene expression in bees (16), whereas the relationship between brain gene expression and brood-feeding behavior in the social insects is currently unknown, and may be weaker than for flight-related behaviors. In bumble bees, workers can transition between specialization on foraging and brood-feeding over the course of a day (40, 41), suggesting great plasticity in these behaviors, which therefore may not involve large-scale changes in brain gene expression (42).

The results of this study also suggest that several candidate genes and molecular pathways previously implicated as important in eusocial insect evolution have also been important in the evolution of sociality in the bumble bees. These include the insulin signaling pathway (12, 33), the transcription factor *kruppel homolog 1* (34, 35), and the inositol signaling pathway (36, 37). The gene *inositol 1,4,5,-tris-phosphate receptor* is a particularly promising candidate for future studies on the shared molecular basis of social evolution in insects, as this gene was differentially expressed in studies

of feeding-related behaviors in *Bombus*, *Polistes* (13), and *Apis* (16), which together represent three independent evolutions of eusociality (17).

In the cross-species comparisons of gene lists derived from *Bombus, Polistes,* and *Apis* microarray studies (14, 16, 18), only the feeding-related lists from *Bombus* and *Polistes* showed a significant degree of overlap. This finding suggests that in regard to brain gene expression, there may be more similarity in the regulation of feeding-related social behaviors relative to reproductive behaviors in *Bombus* and *Polistes*. The greater similarity between the *Bombus* and *Polistes* gene lists relative to *Apis* may also be related to the fact that the *Bombus* and *Polistes* lists were both generated from studies that included different castes (*i.e.,* queens and workers), whereas the *Apis* lists were generated from studies on workers performing different feeding-related tasks (*i.e.,* nurses and foragers). Toth *et al.* (14) also looked for overlap between their *Polistes* feeding- and reproduction-related gene lists and the *Apis* nurse/forager lists examined in this study (16, 18), and found a significant degree of overlap between the *Apis* nurse/forager and *Polistes* feeding-related lists, but not between the *Apis* nurse/forager and *Polistes* reproduction-related lists. In relation to this study, this suggests that there may be greater similarity in the molecular regulation of feeding-related behaviors in *Bombus* and *Polistes*, which are both primitively eusocial, than exists between *Bombus* and *Apis*, which differ more in their degree of social complexity but are more closely related to one another than to *Polistes.*

Together, the results of the analyses included in this study suggest that the multiple independent evolutions of eusociality that occurred in the insects involved a combination of both unique and shared mechanisms. The findings here strengthen the idea that genetic toolkits have played a widespread role in social insect evolution, with feeding- or nutrition-related genetic pathways playing a prominent role.

Acknowledgments

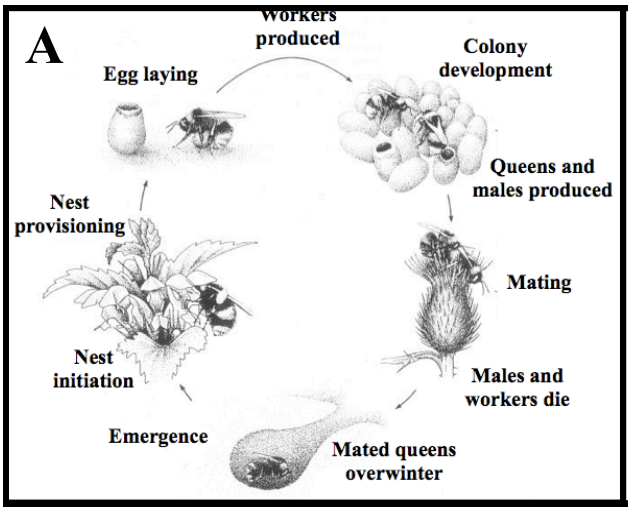
I would like to thank H. Shpigler, Y. Gruber, M. Cohen, and other members of the Bloch lab for comments on experimental design and assistance with collections; A. Toth, T. Daugherty, B. Fischman, and other members of the Robinson lab for comments on the manuscript; M. Band and J. Drnevich for assistance with microarrays; Yad Mordechai for the bees; and K. Varala, M.E. Hudson, and A. Venkat for bioinformatic assistance.

Data Acquisition

The microarray results from this study can be accessed at the ArrayExpress website (http://www.ebi.ac.uk/arrayexpress/).

FIGURES AND TABLES

Figure 8. Experimental design. A) The bumble bee life cycle in temperate climates. Modified with permission from (43). B) 2 x 2 factorial design of experiment.



B

Reproductive	Brood Care +	Brood Care -
+	Foundress	Queen
-	Worker	Gyne

Figure 9. Comparison of *Bombus* and *Polistes* brain gene expression data using all significant probes. A) Heat map derived from HCA from *Bombus* data (*n* = 8044 probes). B) Heat map derived from HCA from *Polistes* (14) data (*n* = 447 probes).

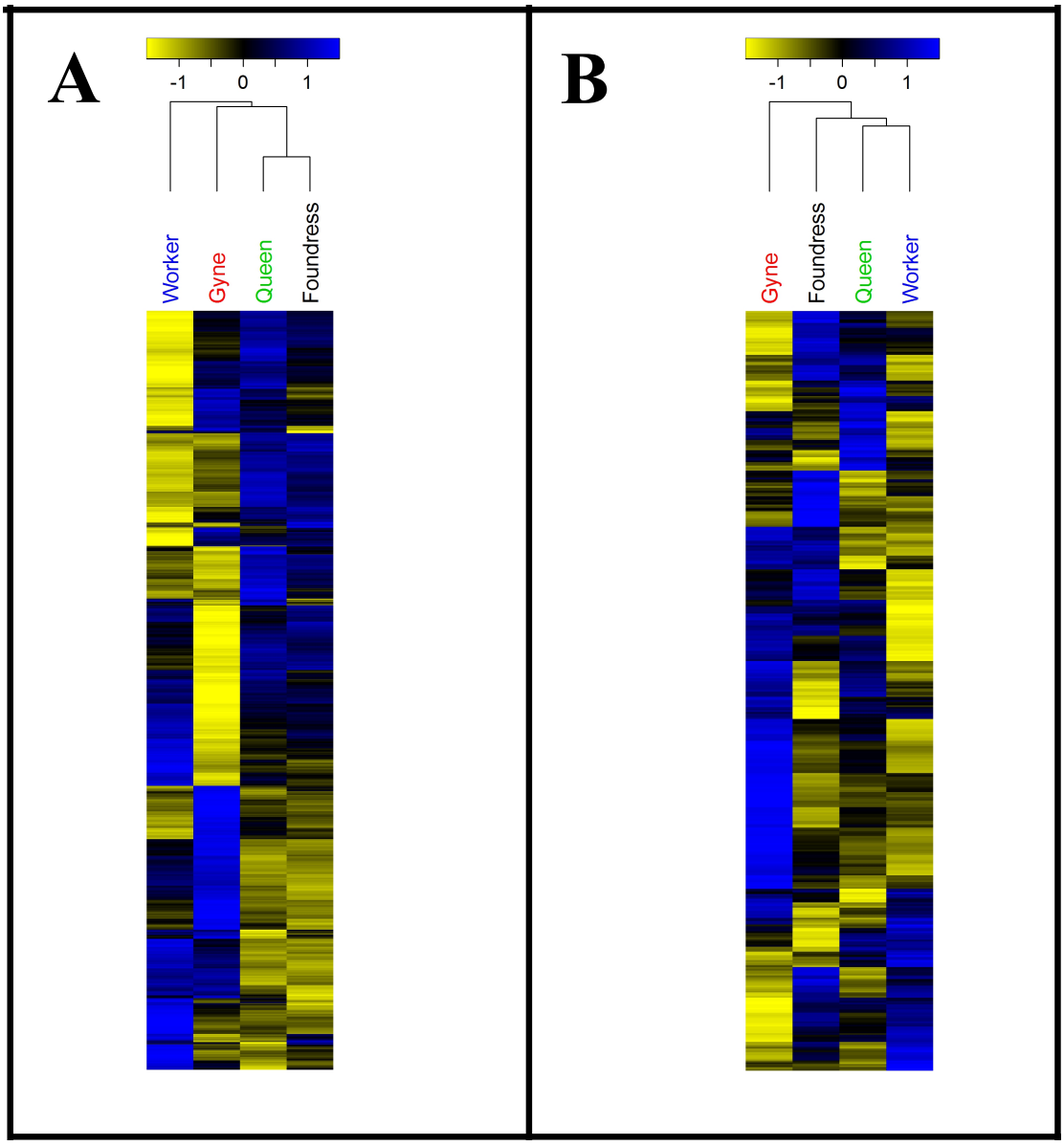


Figure 10. Comparison of *Bombus* and *Polistes* brain gene expression data using subsets of genes related to feeding and reproduction. A) Heat map derived from HCA in Test 1. B) Heat map derived from HCA in Test 2. C) Heat map derived from HCA of *Polistes* qPCR data (13).

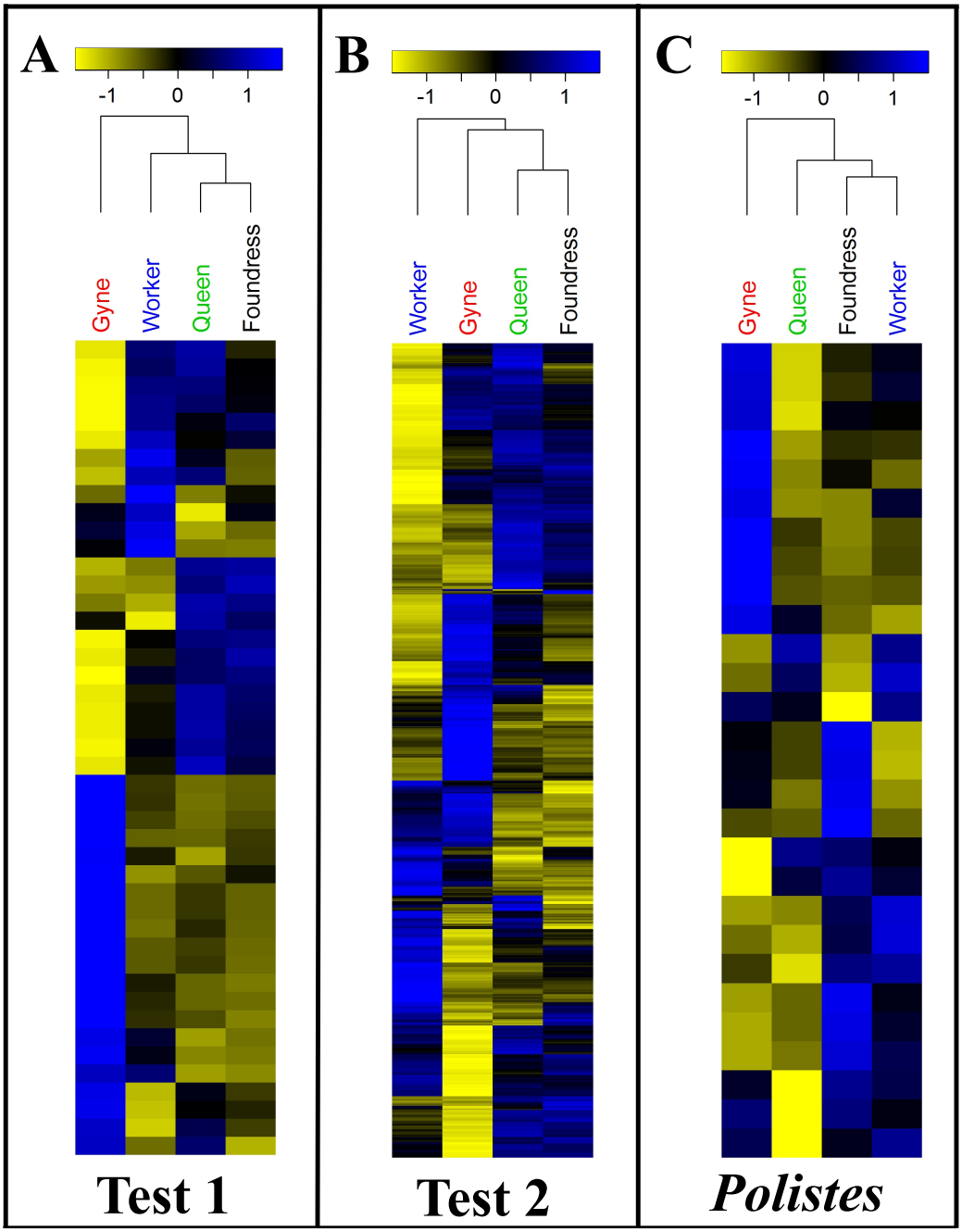


Figure 11. Cross-species comparisons of *Bombus, Polistes,* and *Apis* brain microarray gene lists. Contingency tables of overlap between significant gene lists for *Bombus* Feeding and Reproduction Gene Lists (this study), *Polistes* feeding- and reproduction-related gene lists (14), and *Apis* nurse/forager lists (16, 18). *p* values are from Fisher's Exact Tests (above) and simulations (below). Green = significant ($p < 0.05$, both tests) degree of overlap between lists, white = degree of overlap not significant.

	Bombus Feeding		
Polistes Feeding	$p = 0.022$ $p = 0.009$	Sig	N.S.
	Sig	18	85
	N.S.	136	1190

	Bombus Reproduction		
Polistes Rep.	$p = 0.663$ $p = 0.519$	Sig	N.S.
	Sig	16	24
	N.S.	585	804

	Bombus Feeding		
Apis Nurse/ Forager (cDNA)	$p = 0.222$ $p = 0.187$	Sig	N.S.
	Sig	80	647
	N.S.	93	862

	Bombus Reproduction		
Apis Nurse/ Forager (cDNA)	$p = 0.746$ $p = 0.714$	Sig	N.S.
	Sig	296	431
	N.S.	403	552

	Bombus Feeding		
Apis Nurse/ Forager (oligo)	$p = 0.751$ $p = 0.716$	Sig	N.S.
	Sig	65	528
	N.S.	454	3382

	Bombus Reproduction		
Apis Nurse/ Forager (oligo)	$p = 0.976$ $p = 0.963$	Sig	N.S.
	Sig	203	390
	N.S.	1471	2365

Table 8. Functional terms enriched in the Feeding, Reproduction, and Both Gene Lists. Terms with < 5 genes are not included in table. "GO IDs" are Gene Ontology IDs.

	Term	GO ID	Fold Enrichment	*p* value
Feeding Gene List				
Biological Process Terms				
Behavior	adult locomotory behavior	GO:0008344	2.09	0.02748
	behavior	GO:0007610	1.38	0.03213
	locomotory behavior	GO:0007626	1.51	0.04780
Gene Expression	posttranscriptional gene silencing	GO:0016441	2.15	0.00989
	gene silencing by RNA	GO:0031047	2.15	0.00989
	posttranscriptional gene silencing by RNA	GO:0035194	2.15	0.00989
	dsRNA transport	GO:0033227	2.42	0.01454
	RNA interference	GO:0016246	2.22	0.01784
	RNA transport	GO:0050658	1.82	0.01853
	establishment of RNA localization	GO:0051236	1.82	0.01853
	regulation of gene expression, epigenetic	GO:0040029	1.58	0.02448
	gene silencing	GO:0016458	1.65	0.02547
	negative regulation of gene expression	GO:0010629	1.43	0.03670
	RNA localization	GO:0006403	1.52	0.04943
Ion Transport	metal ion transport	GO:0030001	2.04	0.00100
	di-, tri-valent inorganic cation transport	GO:0015674	2.51	0.01941
	calcium ion transport	GO:0006816	3.14	0.04578
Nucleic Acid Transport	nucleobase, nucleoside, nucleotide and nucleic acid transport	GO:0015931	1.83	0.01399
	nucleic acid transport	GO:0050657	1.82	0.01853
Subcellular Organization	Golgi organization	GO:0007030	2.51	0.01941
Molecular Function Terms				
ATP Binding	ATP binding	GO:0005524	1.24	0.00538
ATPase Activity	ATPase activity, uncoupled	GO:0042624	1.75	0.00094
	ATPase activity	GO:0016887	1.40	0.00717
	ATPase activity, coupled	GO:0042623	1.42	0.01081
GTPase Activity	small GTPase regulator activity	GO:0005083	1.60	0.02173
	Ras GTPase activator activity	GO:0005099	1.79	0.04652
Helicase Activity	ATP-dependent helicase activity	GO:0008026	2.11	0.00170
	purine NTP-dependent helicase activity	GO:0070035	2.11	0.00170
	helicase activity	GO:0004386	1.63	0.01701
	RNA helicase activity	GO:0003724	2.19	0.01915
Ion Binding	metal ion binding	GO:0046872	1.14	0.02049
	ion binding	GO:0043167	1.13	0.02983
	cation binding	GO:0043169	1.12	0.03700
Nucleic Acid Binding	purine ribonucleotide binding	GO:0032555	1.23	0.00216
	ribonucleotide binding	GO:0032553	1.23	0.00216
	purine nucleotide binding	GO:0017076	1.21	0.00302
	adenyl ribonucleotide binding	GO:0032559	1.23	0.00599
	adenyl nucleotide binding	GO:0030554	1.21	0.00969

Table 8 (continued)

	purine nucleoside binding	GO:0001883	1.20	0.01178
	nucleoside binding	GO:0001882	1.19	0.01562
Reproduction Gene List				
Biological Process Terms				
Cellular Movement	localization of cell	GO:0051674	1.36	0.01069
	cell migration	GO:0016477	1.35	0.01817
	cell motion	GO:0006928	1.27	0.01934
	cell motility	GO:0048870	1.33	0.02665
Development	gastrulation	GO:0007369	1.87	0.00968
Ion Transport	metal ion transport	GO:0030001	1.42	0.01840
Reproduction	gamete generation	GO:0007276	1.13	0.04849
Molecular Functions				
ATP Binding	ATP binding	GO:0005524	1.10	0.03791
ATPase Activity	ATPase activity	GO:0016887	1.25	0.00426
	ATPase activity, coupled	GO:0042623	1.24	0.01292
	ATPase activity, uncoupled	GO:0042624	1.30	0.02140
Helicase Activity	purine NTP-dependent helicase activity	GO:0070035	1.42	0.04072
	ATP-dependent helicase activity	GO:0008026	1.42	0.04072
Ion Binding	zinc ion binding	GO:0008270	1.15	0.00270
	metal ion binding	GO:0046872	1.09	0.01200
	cation binding	GO:0043169	1.09	0.01579
	transition metal ion binding	GO:0046914	1.10	0.01960
	ion binding	GO:0043167	1.08	0.02022
Lipid Binding	lipid binding	GO:0008289	1.47	0.00149
	phospholipid binding	GO:0005543	1.60	0.02701
Nucleic Acid Binding	purine nucleotide binding	GO:0017076	1.10	0.01362
	ribonucleotide binding	GO:0032553	1.10	0.01937
	purine ribonucleotide binding	GO:0032555	1.10	0.01937
	nucleotide binding	GO:0000166	1.08	0.03415
	nucleoside binding	GO:0001882	1.10	0.03548
	adenyl nucleotide binding	GO:0030554	1.10	0.03972
	purine nucleoside binding	GO:0001883	1.10	0.03994
	adenyl ribonucleotide binding	GO:0032559	1.10	0.04322
Both Gene List				
Biological Process Terms				
Behavior	behavior	GO:0007610	1.28	0.01843
	adult locomotory behavior	GO:0008344	1.69	0.03452
	locomotory behavior	GO:0007626	1.36	0.03825
	adult behavior	GO:0030534	1.50	0.04094
Gene Expression	translation	GO:0006412	1.29	0.00089
	RNA interference	GO:0016246	1.93	0.00487
	posttranscriptional gene silencing by RNA	GO:0035194	1.78	0.00718
	gene silencing by RNA	GO:0031047	1.78	0.00718
	posttranscriptional gene silencing	GO:0016441	1.78	0.00718
	posttranscriptional regulation of gene expression	GO:0010608	1.37	0.01899
	ribonucleoprotein complex assembly	GO:0022618	1.67	0.02266
	translational elongation	GO:0006414	1.95	0.02418
	RNA transport	GO:0050658	1.51	0.02770

Table 8 (continued)

	nucleic acid transport	GO:0050657	1.51	0.02770
	establishment of RNA localization	GO:0051236	1.51	0.02770
Intracellular Transport	intracellular transport	GO:0046907	1.18	0.03799
Ion Transport	anion transport	GO:0006820	1.82	0.01073
	inorganic anion transport	GO:0015698	2.08	0.03642
Nuclear Export	nuclear export	GO:0051168	1.76	0.03183
Nucleic Acid Transport	nucleobase, nucleoside, nucleotide and nucleic acid transport	GO:0015931	1.49	0.02812
Response to Organic Substance	response to organic substance	GO:0010033	1.43	0.02864
Spliceosome	spliceosome assembly	GO:0000245	1.91	0.04372
Molecular Functions				
Aminopeptidase Activity	aminopeptidase activity	GO:0004177	2.13	0.00965
ATP Binding	ATP binding	GO:0005524	1.12	0.03145
ATPase Activity	ATPase activity, uncoupled	GO:0042624	1.54	0.00038
	ATPase activity	GO:0016887	1.28	0.00551
	ATPase activity, coupled	GO:0042623	1.30	0.00688
Cytoskeleton	structural constituent of cytoskeleton	GO:0005200	1.73	0.02068
Helicase Activity	purine NTP-dependent helicase activity	GO:0070035	1.56	0.01748
	ATP-dependent helicase activity	GO:0008026	1.56	0.01748
Ion Transport	anion transmembrane transporter activity	GO:0008509	1.51	0.02744
Nucleotide Binding	adenyl ribonucleotide binding	GO:0032559	1.12	0.03542
Peptidase Activity	serine-type peptidase activity	GO:0008236	1.46	0.04112
	peptidase activity	GO:0008233	1.20	0.04309
	peptidase activity, acting on L-amino acid peptides	GO:0070011	1.20	0.04718
Ribosomes	structural constituent of ribosome	GO:0003735	1.31	0.00759
Structural Molecule Assembly	structural molecule activity	GO:0005198	1.27	0.00340
Translation	translation elongation factor activity	GO:0003746	1.84	0.02823

Table 9. Summary of brain gene expression data in *Bombus* and *Polistes.* 32 genes are from Toth *et al.*'s (13) study of brain gene expression in *Polistes.* "Prov." = significantly associated with nest provisioning behavior; "Rep." = significantly associated with reproduction; "Feed" = significantly associated with brood-feeding. * = significant (FDR-corrected $p < 0.05$); "-" = not significant; cells left blank when gene was not represented on array used for experiment.

Gene	*A. mellifera* Gene ID	*Polistes* Prov. (13)	*Polistes* Prov. (14)	*Polistes* Rep. (13)	*Polistes* Rep. (14)	*Bombus* Feed.	*Bombus* Rep.
translationally controlled tumor protein	GB16412	*	-	*	-	*	*
monocarboxylate transporter 13-like	GB18699	*	-	*	-	-	-
InR2	N.A.	*		*			
retinoid- and fatty acid-binding glycoprotein	GB11059	*	-	-	-		
delta-1-pyrroline-5-carboxylate synthase-like	GB15049	*	-	-	-		
SH3 domain-binding glutamic acid-rich protein homolog	GB19996	*	-	-	-	*	-
inositol oxygenase-like	GB14956	*		-		*	*
inositol-3-phosphate synthase	GB14423	*	-	-	-	-	*
methylthioribose-1-phosphate isomerase-like	GB12470	*	-	-	-	-	*
CG9005-like	GB19397	*		-			
CG11971-like	GB26541	*		-			
elongation factor 2-like	GB19247	-	*	*	*	*	*
vitellogenin	GB13999	-	*	*	-	-	*
GB10722-like	GB10722	-		*		*	*
insulin-like peptide 2	GB10174	-		*			
InR1	N.A.	-		*			
insulin receptor substrate	GB11037	-	-	-	-	*	-
glycogen phosphorylase	GB11892	-	-	-	-	*	*
CG33293-like	GB13004	-	-	-	-	*	*
period	GB19264	-	-	-	-	*	*
malvolio	GB15139	-		-		*	*
SPARC	GB11432	-	-	-	-	-	*
endopeptidase inhibitor	GB14742	-	-	-	-	-	*
ultraspiracle	GB16648	-	-	-	-	-	*
failed axon connections	GB17380	-	-	-	-	-	*
phosphoinositide-3-kinase	GB11687	-	-	-	-		
protein N-terminal glutamine amidohydrolase	GB12019	-	-	-	-		
carbonic anhydrase 1	GB15888	-	-	-	-		
erk7	GB11031	-		-			

Table 9 (continued)

disintegrin and metalloproteinase domain-containing protein 10-like	GB19179	-		-			
J domain-containing protein-like	GB18301	-	-	-	-	-	-
foraging	GB18394	-	*	-	-	-	-

Table 10. Genes present in *Bombus, Polistes,* and *Apis* brood care-related lists. Gene lists were obtained from this study and studies in *Polistes* (14) and *Apis* (16, 18). "Amel ID" refers to *A. mellifera* gene IDs. *Biological process* terms are for *D. melanogaster* orthologs; not all terms for each gene are listed. An additional 9 genes were excluded from the table due to lack of detailed annotation information. For the columns "Whitfield *et al.* (16)" and "Alaux *et al.* (18)", * = significant (FDR-corrected $p < 0.05$); "-" = not significant; cells left blank when gene was not represented on array used for experiment.

Gene	Amel ID	*Biological Process* Terms	Whitfield *et al.* (16)	Alaux *et al.* (18)
alkaline phosphatase 4	GB18702	epithelial fluid transport		*
ariadne	GB15682	protein catabolic process, proteolysis	*	-
calreticulin	GB17282	behavior, locomotion	*	-
elongation factor 2b	GB19247	mitotic cell cycle, translation	*	*
heat shock protein 83	GB14494	cell surface receptor linked signal transduction, circadian rhythm, neuron differentiation, oogenesis		*
hemolectin	GB15018	polysaccharide metabolic process, melanin metabolic process, immune response		*
imaginal disc growth factor 4	GB11665	imaginal disc development, carbohydrate catabolic process	*	-
inositol 1,4,5,-tris-phosphate receptor	GB11927	ion transport, sensory perception, behavior, mushroom body development, cognition, calcium ion homeostasis	*	-
j domain-containing protein CG6693	GB18305	none	*	-
mitochondrial ribosomal protein S29	GB14023	translation, apoptosis		*
nervana 2	GB15097	ion transport, ribonucleotide metabolic process, glial cell differentiation	-	*
organic anion transporting polypeptide 33Ea	GB19862	ion transport	*	-
probable medium-chain specific acyl-CoA dehydrogenase, mitochondrial	GB16579	lipid catabolic process, oxidation reduction	*	-
protein KIAA0664 homolog	GB10340	translation	*	-
ribosomal protein S8	GB13741	translation	-	*
tetratricopeptide repeat protein 14 homolog	GB19103	none	*	-
thiolester containing protein II	GB12605	endocytosis, immune response	*	-
yippee interacting protein 2	GB12956	lipid catabolic process	*	-

Table 11. Genes present in *Bombus* and *Polistes* reproduction-related lists. Gene lists were obtained from this study and a study in *Polistes* (14). "Amel ID" refers to *A. mellifera* gene IDs. *Biological process* terms are for *D. melanogaster* orthologs; not all biological processes for each gene are listed. An additional 15 genes were excluded from the table due to lack of detailed annotation information.

Gene	Amel ID	*Biological Process* Terms
arginine methyltransferase 7	GB18482	methylation
ariadne	GB15682	protein catabolic process
arrestin 2	GB12766	cell surface receptor linked signal transduction, cognition
dreadlocks	GB12942	cell morphogenesis, cell surface receptor linked signal transduction, axonogenesis, insulin receptor signaling pathway, neuron differentiation, neuron development
elongation factor 2b	GB19247	translation, cell cycle
endophilin A	GB14809	synaptic transmission
excitatory amino acid transporter 1	GB30125	synaptic transmission, aging
excitatory amino acid transporter 2	GB16377	organic acid transport
glycerol 3 phosphate dehydrogenase	GB10776	behavior, carbohydrate catabolic process, glycerolipid metabolic process
glycerol kinase	GB13023	glycerol metabolic process
heat shock factor	GB13400	transcription, response to temperature stimulus
heterogeneous nuclear ribonucleoprotein at 87F	GB11598	regulation of RNA splicing
larval serum protein 2	GB30362	none
meiotic central spindle	GB17567	none
nidogen/entactin	GB14424	cell adhesion
organic cation transporter	GB13312	ion transport, apoptosis
pyruvate kinase	GB10695	glucose metabolic process
receptor-mediated endocytosis protein 6 homolog	GB12354	regulation of protein localization
ribosomal protein S28b	GB13692	translation
thiolester containing protein II	GB12605	immune response

LITERATURE CITED

1. Wilson EO (1971) *The Insect Societies.* Harvard Univ. Press.
2. Michener CD (1969) Comparative social behavior of bees. *Ann Rev Entomol* 14:299-342.
3. Evans HE, West-Eberhard MJ (1970) *The Wasps* (Univ. of Michigan Press, Ann Arbor, MI).
4. West Eberhard MJ (1987) In *Animal Societies: Theories and Facts,* Ito Y, Brown JL, Kikkawa J. Eds. Japan Sci. Soc. Press, Tokyo, pp 35-51.
5. West Eberhard MJ (1996) In *Natural History and Evolution of Paper Wasps*, Turillazzi S, West Eberhard MJ, Eds. Oxford Univ. Press, New York, pp 290–317.
6. Linksvayer T, Wade M (2005) The evolutionary origin and elaboration of sociality in the aculeate Hymenoptera: maternal effects, sib-social effects, and heterochrony. *Quart Rev Biol* 80:317-336.
7. Toth AL, Robinson GE (2007) Evo-devo and the evolution of social behavior. *Trends Genet* 23:334-341.
8. Tautz D, Domazet-Loso T (2011) The evolutionary origin of orphan genes. *Nat Rev Genet* 12:692-702.
9. Amdam G, Csondes A, Fondrk M, Page R (2006) Complex social behaviour derived from maternal reproductive traits. *Nature* 439:76–78.
10. Wilson EO & Hölldobler B (2005) Eusociality: origin and consequences. *Proc Natl Acad Sci USA* 102: 13367–13371.
11. Woodard SH, *et al*. (2010) Genes involved in convergent evolution of eusociality in bees. *Proc Natl Acad Sci USA* 7472-7477.
12. Smith CR, Toth AL, Suarez AV, Robinson GE (2010) Genetic and genomic analyses of the division of labour in insect societies. *Nat Rev Genet* 9:735-748.
13. Toth AL, *et al*. (2007) Wasp gene expression supports an evolutionary link between maternal behavior and eusociality. *Science* 318: 441-444.

14. Toth AL, Varala K, Henshaw MT, Rodriguez-Zas SL, Hudson ME, Robinson GE (2010) Brain transcriptomic analysis in paper wasps identifies genes associated with behaviour across social insect lineages. *Proc Roy Soc Lond B* 277:2139-2148.
15. Free JB, Butler CG (1959) *Bumblebees*. Collins, London.
16. Whitfield CW, Cziko AM, Robinson GE (2003) Gene expression profiles in the brain predict behaviour in individual honey bees. *Science* 302:296–299.
17. Cameron SA, Mardulyn P (2001) Multiple molecular data sets suggest independent origins of highly eusocial behavior in bees (Hymenoptera:Apinae). *Syst Biol* 50:194-214.
18. Alaux C, *et al*. (2009). Honey bee aggression supports a link between gene regulation and behavioral evolution. *Proc Natl Acad Sci USA* 106:15400-15405.
19. Bloch G, Borst DW, Huang ZY, Robinson GE, Hefetz A (1996) Effects of social conditions on Juvenile Hormone mediated reproductive development in *Bombus terrestris* workers. *Physiol Entom* 21:257-267.
20. Duchateau MJ, Velthuis HHW (1989) Ovarian development and egg-laying in workers of *Bombus terrestris*. *Entom Experiment Et Applic* 51:199-213.
21. Röseler PF, van Honk CGJ (1990) Castes and reproduction in bumblebees. In: *Social Insects: an Evolutionary Approach to Castes and Reproduction,* Engels W, Ed. Springer, Berlin, Heidelberg, New York. pp 147–166.
22. Schulz DJ, Robinson GE (1999) Biogenic amines and division of labor in honey bee colonies: Behaviorally related changes in the antennal lobes and age-related changes in the mushroom bodies. *J Comp Physiol A* 184:481-488.
23. Sen Sarma M, Rodriguez-Zas S, Hong F, Zhong S, Robinson GE (2009) Transcriptomic profiling of central nervous system regions in three species of honey bee during dance communication behavior. *PLoS One* 4:e6408.

24. R Development Core Team (2011) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.
25. Smyth GK (2005) Limma: linear models for microarray data. In: Bioinformatics and Computational Biology Solutions using R and Bioconductor, R. Gentleman, V. Carey, S. Dudoit, R. Irizarry, W. Huber (eds.), Springer, New York, pages 397-420.
26. Smyth GK, Speed TP (2003) Normalization of cDNA microarray data. *Methods* 31:265-273.
27. Smyth GK (2004) Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. *Stat App Genetics Mol Biol* 3: Article 3.
28. Smyth GK, Michaud J, Scott H (2005) The use of within-array replicate spots for assessing differential expression in microarray experiments. *Bioinformatics* 21:2067-2075.
29. Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J Royal Stat Soc* B 57:289-300.
30. HBGSC (2006) Insights into social insects from the genome of the honeybee *Apis mellifera*. *Nature* 443:931-949.
31. Huang DW, Sherman BT, Lempicki RA (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protoc* 4:44-57.
32. Warnes GR Includes R source code and/or documentation contributed by: Ben Bolker (in alphabetical order), Lodewijk Bonebakker, Robert Gentleman, Wolfgang Huber Andy Liaw, Thomas Lumley, Martin Maechler, Arni Magnusson, Steffen Moeller, Marc Schwartz and Bill Venables (2010). gplots: Various R programming tools for plotting data. R package version 2.8.0.
33. Ament SA, Corona M, Pollock HS, Robinson GE (2008) Insulin signaling is involved in the regulation of worker division of labor in honey bee colonies. *Proc Natl Acad Sci USA* 105:4226-4231.

34. Grozinger CM, Sharabash NM, Whitfield CW, Robinson GE (2003) Pheromone-mediated gene expression in the honey bee brain. *Proc Natl Acad Sci USA* 100:14519-14525.
35. Shpigler H, Patch HM, Cohen M, Fan Y, Grozinger CM, Bloch G (2010) The transcription factor *Kruppel homolog 1* is linked to hormone mediated social organization in bees. *BMC Evol Biol* 10:120.
36. Agrawal N, Padmanabhan N, Hasan G (2009) Inositol 1,4,5-trisphosphate receptor function in *Drosophila* insulin producing cells. *Plos One* 4:e6652.
37. Agrawal N, Venkiteswaran G, Sadaf S, Padmanabhan N, Banerjee S, Hasan G (2010) Inositol 1,4,5-trisphosphate receptor and dSTIM function in *Drosophila* insulin-producing neurons regulates systemic intracellular calcium homeostasis and flight. *J Neuroscience* 30:1301-1313.
38. Toth AL, Bilof KBJ, Henshaw MT, Hunt JH, Robinson GE (2009) Lipid stores, ovary development, and brain gene expression in *Polistes metricus* females. *Insectes Soc* 56:77-84.
39. Reeve HK (1991) *Polistes*. In: The Social Biology of Wasps (Ross K.G. and Matthews R.G., Eds). Cornell University Press, Ithaca, NewYork. pp 99 – 148.
40. Brian AD (1952) Division of labour and foraging in *Bombus agrorum* Fabricius. J Animal Ecol 21:223-240.
41. Free JB (1955) The division of labour within bumblebee colonies. *Insectes Soc* 2:195-212.
42. Burmeister SS, Jarvis ED, Fernald RD (2005) Rapid behavioral and genomic responses to social opportunity. *PloS Biol* 3:e363.
43. Prys-Jones OE, Corbet SA (1987) *Bumblebees*. Cambridge University Press.

APPENDIX A

Table A1. Brood-feeding behavior and colony development data.

#	Workers	Stage	Feeding, Day 3	Feeding, Day 5	Feeding, Day 7	Feeding, Total	No. Eggs	No. Larvae	No. Pupae
1	-	Late	8	14	16	38	18	38	10
2	-	Late	6	2	7	15	0	42	8
3	-	Late	6	6	5	17	0	63	12
4	-	Late	3	4	6	13	31	24	14
5	-	Late	7	5	4	16	31	14	14
6	-	Late	2	5	1	8	0	66	11
7	-	Late	3	4	1	8	25	58	13
8	-	Late	0	3	5	8	24	48	12
9	-	Late	8	6	5	19	0	42	10
10	-	Late	3	8	1	12	28	34	13
11	-	Late	6	6	4	16	41	28	23
12	-	Late	5	4	11	20	9	39	7
13	+	Late	1	4	0	5	78	71	24
14	+	Late	3	7	1	11	56	49	21
15	+	Late	1	1	0	2	56	50	13
16	+	Late	0	0	0	0	93	12	27
17	+	Late	4	0	2	6	64	50	14
18	+	Late	5	0	0	5	86	32	19
19	+	Late	2	1	4	7	96	51	23
20	+	Late	5	2	2	9	47	36	16
21	+	Late	4	0	0	4	155	10	4
22	+	Late	4	0	0	4	77	67	22
23	+	Late	0	0	0	0	36	47	22
24	-	Early	0	8	5	13	7	5	5
25	-	Early	4	17	9	30	24	7	10
26	-	Early	2	4	3	9	38	4	5
27	-	Early	5	5	6	16	8	5	7
28	-	Early	0	6	11	17	0	13	0
29	-	Early	11	14	12	37	19	8	10
30	-	Early	6	11	5	22	13	6	12
31	-	Early	6	9	10	25	8	9	6
32	-	Early	6	7	3	16	22	13	12
33	-	Early	1	3	2	6	0	8	0
34	-	Early	8	8	12	28	8	11	9
35	+	Early	1	0	0	1	27	6	0
36	+	Early	0	2	0	2	19	7	0
37	+	Early	0	2	3	5	41	4	0
38	+	Early	1	0	0	1	87	12	0
39	+	Early	1	0	0	1	21	10	3
40	+	Early	0	0	1	1	47	37	10
41	+	Early	0	1	3	4	73	27	15

Table A1 (continued)

42	+	Early	6	7	3	16	25	20	15
43	+	Early	2	4	0	6	20	25	6
44	+	Early	1	2	0	3	39	3	0
45	+	Early	2	0	2	4	10	7	0

Table A2. Descriptive statistics of brood-feeding behavior and colony development data.

	Group	Sample Size	Mean	Standard Error
Brood-feeding, Day 3	EQ + W	11	1.27	0.52
	EQ - W	11	4.45	1.05
	LQ + W	11	2.64	0.58
	LQ - W	12	4.75	0.73
Brood-feeding, Day 5	EQ + W	11	1.64	0.66
	EQ - W	11	8.36	1.28
	LQ + W	11	1.36	0.68
	LQ - W	12	5.58	0.89
Brood-feeding, Day 7	EQ + W	11	1.09	0.41
	EQ - W	11	7.09	1.15
	LQ + W	11	0.82	0.40
	LQ - W	12	5.50	1.26
Brood-feeding, Total	EQ + W	11	4.00	1.31
	EQ - W	11	19.91	2.84
	LQ + W	11	4.82	1.03
	LQ - W	12	15.83	2.35
Egg-laying	EQ + W	11	37.18	7.23
	EQ - W	11	13.36	3.45
	LQ + W	11	76.73	9.73
	LQ - W	12	17.25	4.28
Larvae	EQ + W	11	14.36	3.38
	EQ - W	11	8.09	0.95
	LQ + W	11	43.18	5.87
	LQ - W	12	41.33	4.52
Pupae	EQ + W	11	4.45	1.85
	EQ - W	11	6.91	1.28
	LQ + W	11	18.64	1.96
	LQ - W	12	12.25	1.17

Table A3. *p* values from pairwise t-tests of brood-feeding behavior and colony development data from four groups. Values listed in table are *p*-values from 2-tailed t-tests. Boxes shaded in blue are significant at $p < 0.05$.

		EQ - W	LQ + W	LQ - W
Feeding, Day 3	**EQ + W**	0.01329	0.09530	0.00102
	EQ - W		0.14402	0.81671
	LQ + W			0.03564
	LQ - W			
Feeding, Day 5	**EQ + W**	0.00015	0.77680	0.00214
	EQ - W		0.00010	0.08518
	LQ + W			0.00128
	LQ - W			
Feeding, Day 7	**EQ + W**	0.00008	0.64122	0.00424
	EQ - W		0.00005	0.36375
	LQ + W			0.00261
	LQ - W			
Feeding, Total	**EQ + W**	0.00006	0.62999	0.00033
	EQ - W		0.00007	0.27834
	LQ + W			0.00044
	LQ - W			
Egg-Laying	**EQ + W**	0.00755	0.00391	0.02471
	EQ - W		0.00001	0.49263
	LQ + W			0.00001
	LQ - W			
Larvae	**EQ + W**	0.08911	0.00039	0.00012
	EQ - W		0.00001	0.00000
	LQ + W			0.80357
	LQ - W			
Pupae	**EQ + W**	0.28776	0.00004	0.00158
	EQ - W		0.00007	0.00550
	LQ + W			0.00937
	LQ - W			